



Residential Segregation in 5 European Countries  
Technical Report

Michael Meinild Nielsen, Karen Haandrikman, Henning Christiansen, Rafael Costa, Bart Sleutjes, Adrian F. Rogne, and Marcin Stonawski

**ResSegr Working Paper 2017:2**

*© Copyright is held by the author(s). ResSegr Working Papers receive only limited review. Views and opinions expressed in ResSegr Working Papers are attributable to the authors and do not necessarily reflect those held at the institutions the authors are affiliated with.*

# **Residential Segregation in 5 European Countries - Technical Report**

*Michael M. Nielsen*

Stockholm University, Department of Human Geography

*Karen Haandrikman*

Stockholm University, Department of Human Geography

*Henning Christiansen*

Statistics Denmark

*Rafael Costa*

Vrije Universiteit Brussel, Department of Sociology, Faculty of Economic, Social and Political,  
Sciences and Solvay Business School

*Bart Sleutjes*

Netherlands Interdisciplinary Demographic Institute (NIDI)/KNAW/ University of Groningen

*Adrian F. Rogne*

Department of Sociology and Human Geography, University of Oslo

*Marcin Stonawski*

Department of Sociology and Human Geography, University of Oslo

## Preface

The research project “Residential segregation in in five European countries – A comparative study using individualized scalable neighbourhoods” (ResSegr) started in August 2014 as a cooperation between researchers at Stockholm University (Department of Human Geography), the University of Oslo (Department of Sociology and Human Geography), Statistics Denmark, the Netherlands Interdisciplinary Demographic Institute and the Vrije Universiteit Brussels (Interface Demography). Funding was granted by the Joint Programme Initiative Urban Europe.

This is the technical report documenting the processes that have led to the making of the harmonized multi-country datasets with segregation indicators that was one of the main outputs of the project. In the report, we provide a description of the national datasets and the geographical coordinates, the definition of indicators and a description of the software used to produce the data. Similarities as well as differences between the different national datasets and indicators are highlighted. One chapter pays attention to the various ethical and privacy considerations that were considered in the creation of the dataset so that privacy of individuals could be protected. More information about the project can be found at [www.residentialsegregation.org](http://www.residentialsegregation.org).

Karen Haandrikman, project leader

November 2017

## Contents

PREFACE.....	3
1. INTRODUCTION.....	7
2. CONSTRUCTION OF THE HARMONIZED MULTI-COUNTRY DATASETS.....	10
3. DESCRIPTION OF THE 13 VARIABLES/INDICATORS .....	12
4. GENERAL DESCRIPTIONS OF THE DATA USED FOR EACH COUNTRY .....	15
4.1 Denmark.....	15
4.2 Sweden.....	16
4.3 Norway .....	17
4.4 Netherlands.....	19
4.5 Belgium .....	21
5. THE GEOGRAPHICAL INFORMATION .....	24
5.1 Denmark.....	24
5.2 Sweden.....	24
5.3 Norway.....	25
5.4 Netherlands.....	26
5.5 Belgium .....	26
6. DETAILED DESCRIPTION OF VARIABLES.....	28
6.1 Indicator 1a. Share of persons aged 25 and above who have a personal disposable income below 60 % of the median level .....	28
6.2 Indicator 2a. Immigrants' share of the population.....	31
6.3 Indicator 2b. EU28/EFTA Immigrants' share of the population .....	34
6.4 Indicator 2c. NON-EU28/EFTA Immigrants' share of the population .....	38
6.5 Indicator 2d. Share of the population for persons with foreign origin (immigrants+descendent) .....	43

6.6	Indicator 2e. Share of the population for persons with foreign origin (immigrants+descendent) in an EU28/EFTA country .....	47
6.7	Indicator 2f. Share of the population for persons with foreign origin (immigrants+descendent) in a NON-EU28/EFTA country .....	53
6.8	Indicator 3a. Share of persons 25-64 year olds who have completed tertiary education .....	58
6.9	Indicator 3b. Share of persons among native born 25-64 year olds who have completed tertiary education .....	62
6.10	Indicator 4a. Share of 25-64 year olds who have a level of taxable earned income in the highest decile .	67
6.11	Indicator 5a. Share of persons aged 18-64 year old who received social assistance at some point in the reference year .....	70
6.12	Indicator 6a. Share of 25-64 year olds in employment.....	73
6.13	Indicator 6b. Share of 30-59 year old in employment .....	76
7.	SUMMARY OF DIFFERENCES BETWEEN VARIABLES AND COUNTRIES FOR DIFFERENT YEARS ..	80
7.1	General remarks about the data in the five countries.....	80
7.2	Geographic information.....	80
7.3	Immigrants and descendants.....	81
7.4	Tertiary education .....	81
7.5	Income.....	82
7.6	Social assistance .....	83
7.7	Employment .....	83
8.	ETHICAL CONSIDERATIONS AND PRIVACY.....	85
8.1.	Denmark .....	85
8.2.	Sweden .....	86
8.3.	Norway.....	87
8.4.	Netherlands.....	88
8.5.	Belgium .....	89
	REFERENCES.....	91



## 1. Introduction

Residential segregation has for some time been a hot topic among planners, human geographers and sociologists. Defined as the degree to which two or more groups live separately from one another in different parts of the (urban) environment (Massey and Denton 1988: 282), residential segregation is considered to be a substantial threat to social inclusion and the welfare state both in Europe and beyond (Finney and Catney 2012; Lichter et al. 2012; Malmberg et al. 2013; Olzak et al. 1996; Schierup and Ålund 2011). However, our current understanding of segregation is hindered by a lack of comparative studies on segregation levels, making it nearly impossible to test theories, analyse factors that contribute to segregation, examine effects of segregation and to suggest interventions to address problems associated with high levels of segregation.

The ResSegr research project proposes an innovative measure of segregation, where neighbourhoods are defined from around individuals instead of being based on administrative borders. Using this methodology, we will be the first to present comparable segregation measures, both over time and across urban areas and countries. The project focuses on socio-demographic segregation, such as segregation by education, high income, unemployment, poverty, social assistance, country of birth and immigrant background. Mapping the variation in geographical contexts in the urban areas in North-western Europe using comparable ethnic and socio-economic indicators will open up new possibilities for addressing questions of central importance for urban analysis and urban policies.

ResSegr has four goals:

1. analysing the patterns of segregation;
2. analysing the determinants of segregation;
3. analysing the effects of segregation;
4. examining policy implications regarding the effectiveness of interventions such as area-based programmes.

This technical report describes the first step in the ResSegr research project: the computation of measures of residential segregation based on individualized scalable neighbourhoods for five European countries, assembled in harmonized multicountry datasets. Using the indicators in de

the dataset, we will be able to analyse patterns of segregation both over time and between different geographical areas. This makes it possible to study whether segregation patterns are similar or different across countries and urban areas. Subsequently, different theories about what the causes of residential segregation are can be evaluated. What role do, for example, income inequality, ethnic composition, housing policies, and different planning ideologies play for the establishment of different patterns of residential segregation? In addition, such comparative data enable studies on the effects of residential segregation in places with different political and economic systems. Comparisons of contextual effects on social inclusion with outcomes such as education, employment, poverty, health, and crime will be topic of research in a later stage of the project. This will in its very essence give policy implications and indicate the necessity for evaluate planning policies in different sectors of society; health, schooling, housing and policies of income distribution. It will thus be possible to examine policy implications regarding the effectiveness of measures such as area-based programmes.

Previous approaches to studying segregation are hampered by three main problems. First, many of the existing studies use the dissimilarity index, but it is generally acknowledged that this index is not invariant to the differences in areal subdivisions between countries and that it is poorly linked to theories of segregation and segregation effects (Duncan and Duncan 1955; Morrill 1991). Second, segregation studies generally use aggregate values for administratively defined areas. However, geographical units often differ in size, function and distribution between regions and over time, known as the Modifiable Area Unit Problem (MAUP) (Openshaw 1984). MAUP affects all quantitative analyses making use of geographical delineations, but in comparative studies, the problem is aggravated, as spatial units that represent data tend to differ structurally between regions and over time. In addition, the differences in population data systems between countries with different types of data and different output geographies make comparisons difficult (Lloyd and Shuttleworth 2009). Third, how a neighbourhood affects an individual may operate at different scales. Over the life course, the size of the neighbourhood relevant to the individual may increase, and depending on the type of segregation, different scales may apply. Summarizing, the lack of standardised measures and comparative data has hindered progress in segregation studies.

In recent years, there have been attempts to provide a more geographically inclusive approach to segregation research. American segregation researchers have started to make a distinction

along geographical levels, and between micro and macro segregation (Fisher et al. 2004; Lichter et al. 2015a; Reardon et al. 2008). However, most studies still use administratively defined areas. Others have modified segregation and isolation indices by introducing spatially weighted matrices to reflect the extent of contact between spatial units (e.g. Wong 2004).

We utilize an innovative methodology that creates individualized neighbourhoods. A few recent studies have used similar type of neighbourhoods, which have also been called bespoke and egocentric neighbourhoods (Chaix et al. 2005; 2009; Clark et al. 2015; Fowler 2015; Lee et al. 2008; MacAllister et al. 2001; Malmberg et al. 2011; Östh et al. 2014; 2015; Reardon et al. 2008; 2009; Shuttleworth and Östh 2013). Using such a multiscale approach, problems related to using aggregate values for administratively defined areas such as the MAUP, are circumvented. The main advantage of using neighbourhoods with a pre-determined number of nearest neighbours is that it can be applied in the same way in different national contexts. When studies use individualized neighbourhoods with the same number of neighbours, the resulting measures of residential segregation and its effects will be exactly comparable.

The harmonized multi-country datasets are based on 13 different indicators measured on nine scale levels. In chapters 5 to 7, the indicators are discussed. The scale levels vary with the smallest being 200 nearest neighbours and the largest being 51,200 nearest neighbours, in order to differentiate between small-scale and large-scale neighbourhoods. Taking the example of the indicator employment, the harmonized multi-country datasets include indicators measuring the share of individuals that are employed among the nearest 200 to 51,200 persons for each populated location in the different countries. The inclusion of different scales is one of the main strengths of as it makes it possible to study segregation at different scales. The construction of the data is described in chapter 2.

## 2. Construction of the harmonized multi-country datasets

The harmonized multi-country datasets consist of 13 indicators, described in chapters 5 to 7, calculated for a range of individualized scalable neighbourhoods. The methodology is a Geographical Information Systems (GIS)-based approach, and uses the increased availability of geocoded individual data to construct individualized scalable neighbourhoods. This means that neighbourhoods are defined from around individuals instead of being based on administrative borders.

Technically, we expand a geographical buffer around the location of each individual until this buffer contains a pre-determined number of nearest neighbours. This number of neighbours is then used to compute aggregate statistics such as, for example, the share of individuals with higher education. By varying the number of nearest neighbours that are included in the buffer it also becomes possible to measure the population compositions of these individualized neighbourhoods at different scale levels. If only the 200 nearest neighbours are included one gets a measure of the closest environment. If the buffer is expanded to include the 800 nearest neighbours one gets a measure that corresponds to a more extended neighbourhood, and by extending the buffer to include 6400 neighbours further one can get a measure of more large-scale environments. The method is based on the distance as-the-crow-flies, and thus does not take physical barriers, national borders and other political or natural barriers into account.

The specialized software for constructing individualized scalable neighbourhoods that was used in the project is called EquiPop, which has been developed by John Östh at Uppsala University. On the EquiPop webpages, one can download the software for free. There is also a manual for the software available online, as well as several peer-reviewed articles using the software: <http://equipop.kultgeog.uu.se/index.html>

Each of the five countries used the software EquiPop to produce the input for the harmonized multi-country datasets. The different scale levels that were used were the 200, 400, 800, 1600, 3200, 6400, 12,800, 25,600, and 51,200 nearest persons for all populated locations in each country.

The EquiPop software input data consists of coordinates in a regular grid and total population and group population for each grid. For example, if one wants to calculate individualized

neighbourhoods measuring the ratio of employed people among the 200 nearest 25-64 year olds, the input data would consist of:

1. Id column: id number for the users convenience
2. EastWest column: X coordinate
3. NorthSouth column: Y coordinate
4. CountAll column: Number of 25-64 year olds in each grid
5. CountSubGroup column: Number of 25-64 year olds that are employed in each grid

The EquiPop software then lets you specify the size of the individualized neighbourhood. In the current example the size then would be 200, representing the 200 nearest 25-64 year olds. One also has the option to calculate the distance that is needed to for each populated location to include the nearest 200 nearest 25-64 year olds.

The output data from EquiPop, using the example of employed among the 200 nearest 25-64 year, includes information of the actual size of the individualized neighbourhood. The size is always at least the size set for the individualized neighbourhood, but quite often it is a bit higher due to nature of aggregating the data into grids. The output data also includes the ratio of 25-64 year olds that are employed in the individualized neighbourhood. See the EquiPop manual for a more detailed description of the input and output data and the different options available in the software (Östh 2014).

To not be able to connect individuals to any of the ratios 1 and 0, these were re-coded to either 0.95 and 0.05 or 0.9 and 0.1, depending on the country (see chapter 8).

### **3. Description of the 13 variables/indicators**

At the start of the project, we decided that the harmonized multi-country datasets would contain indicators on six socio-demographic topics: poverty, foreign-born, education, high income, social assistance and employment. Below, the indicators that were commonly decided in ResSegr are defined.

#### **Poverty**

In defining poverty, we base our indicator on the Eurostat definition of the at-risk-of-poverty rate, defined as the share of people with an equivalised disposable income below the at-risk-of-poverty threshold, which is set at 60 % of the national median equivalised disposable income after social transfers. It would be preferred to use disposable income for families, as well as information on family size and composition. However, most countries in the project do not have data on these aspects. In order to increase comparability, we decided to define income as disposable income by individual persons. A second change compared to the Eurostat definition is that the ResSegr indicator is restricted to persons aged 25 and above.

*Indicator 1a. Share of persons aged 25 and above who has a personal disposable income below 60 % of the median level.*

#### **Foreign-born**

To be able to distinguish between persons who are foreign-born and persons with foreign origin, two sets of indicators were decided upon. Information on country of origin is the basis for grouping persons by foreign-born and foreign-origin.

Immigrants are persons who are foreign born and who have no native-born parents. If no information exists on the birthplace of the parents the person should still be considered an immigrant if he/she is foreign-born.

A descendant is a person who is native-born but who has no parents who are native born. If one parent is native-born the person is not a descendant. If no information exists on the birthplace of the parents the person should not be considered a descendant if he/she is native-born.

Persons of foreign origin are those that are foreign-born plus descendants.

For immigrants, the country of origin is the person's own country of birth. For descendants, the origin country of the parents is defined as the mother's country of birth. If no information exists on the mother's country of birth, the father's country of birth is taken instead.

Countries are grouped into two groups defined by country of origin. The first group consists of persons who are born in the EU28 or EFTA. The EFTA countries are Norway, Switzerland, Iceland and Liechtenstein. The second group includes all remaining countries.

*Indicator 2a. Immigrants' share of the population*

*Indicator 2b. EU/EFTA Immigrants' share of the population*

*Indicator 2c. NON-EU/EFTA Immigrants' share of the population*

*Indicator 2d. Share of the population for persons with foreign origin (immigrants+descendants)*

*Indicator 2e. Share of the population for persons with foreign origin (immigrants+descendants) in an EU/EFTA country*

*Indicator 2f. Share of the population for persons with foreign origin (immigrants+descendants) in a NON-EU/EFTA country*

## **Education**

In the project we have a strong interest in the share of persons with tertiary education. The appropriate age group for this indicator was decided as people aged 25-64.

Tertiary education may be defined using ISCED codes (International Standard Classification of Education), in this case the ISCED1997 version. According to this classification, tertiary education comprises the levels 5 and 6.

Several countries have problems registering educational levels for persons who obtained their education abroad. This leads to some problems in educational levels for foreign-born. We therefore decided to have two different indicators, one for the whole population, and one for persons that are native born.

*Indicator 3a. Share of persons 25-64 year olds who have completed tertiary education*

*Indicator 3b. Share of persons among native born 25-64 year olds who have completed tertiary education*

## **High income**

Similar to the data used for indicator 1a, the same problem arises that we only have data on personal income for all five countries in the project.

Income is defined as taxable earned income, that is, only income from wages or net earnings from self-employment are included.

The age group is restricted to 25-64 year olds.

Persons are ranked according to taxable earned income and grouped into deciles. Persons who are in the highest income decile are defined as high income earners.

*Indicator 4a. Share of 25-64 year olds who have a level of taxable earned income in the highest decile*

## **Social assistance**

In the countries included in the project, social assistance is organized in different ways. To increase comparability, we define social assistance as assistance given to persons with economic problems who have no other means to support their living.

The age group is restricted to 18-64 old persons.

*Indicator 5a. Share of persons aged 18-64 year old who received social assistance at some point in the reference year*

## **Employment**

We base our definition of employed persons on the definition of the International Labour Organization: all persons above a specified age who during a specified brief period, either one week or one day, had paid employment or were self-employed.

Any employment during the period is counted as employment.

Two indicators are defined with varying age-groups.

*Indicator 6a. Share of 25-64 year olds in employment*

*Indicator 6b. Share of 30-59 year old in employment*

## **4. General descriptions of the data used for each country**

### *4.1 Denmark*

All statistical registers used for the indicators in the ResSegr project are administered by Statistics Denmark.

#### **Geocoded data**

The registration of the local residential population in the municipality is done by the local registers in each of the 98 municipalities. Address coordinates are registered by the Danish municipalities and are published in Danish registers like AWS and BBR (dwelling and building register).

#### **Foreign-born**

Information about country of birth of the person and the parents comes from the Central Population Register (CPR) run by the Ministry of Social and Interior. CPR links parents with their (adopted) children. The information is available in CPR back in time until 1968 when CPR was established in Denmark.

#### **Education**

Education information is based on the student register which is a longitudinal register which follows the educational careers of registered persons. It is updated with annual reports from the Danish educational institutions.

#### **Income**

The primary source of income data is administrative data from the Danish tax authorities including the final tax statements and the e-income register. In addition, registers from the municipalities and unemployment funds are used to identify the exact types of transfers persons received. The population register provides background information on demographics. The wage register provides information on line of business, working hours and ISCO codes. The Tax real-estate evaluation register is used for estimating imputed rent.

## **Social assistance**

The indicator on social assistance is based on data from the Cash Benefits register that is administered and maintained by Statistics Denmark. The Cash Benefits register is a census and is based on information from administrative IT-systems in the Danish municipalities.

## **Employment**

The source for the employment indicator is the register-based labour force statistics, which contains information on labour market affiliation for the population at the end of November each year.

### *4.2 Sweden*

The Swedish register data is administered in the GeoStar database, which is a collection of databases managed at Statistics Sweden, and accessed via the MONA system (<http://www.scb.se/mona-en/>).

## **Geocoded data**

Information about geographical coordinates are based on Geografidatabasen. Geografidatabasen contains annual geographical information connected to properties, with coordinates collected by the Swedish mapping, cadastral and land registration authority (Statistics Sweden 2011b).

## **Foreign-born**

Information about (1) age and (2) country of birth is based on Statistics Sweden's register of the total population (Registret över Totalbefolkningen). The register of the total population is an extract from the Swedish Tax Agency population register (Folkbokföringsregistret) (Statistics Sweden 2011b).

Information about who are a person's parents is collected through the Swedish multi-generation register (Flergenerationsregistret). The multi-generation register is based on the Swedish Tax Agency population register and is a register of persons who have been registered in

Sweden at some time since 1961 and who were born in 1932 or later – so-called index persons. The register contains connections between index persons and their biological parents. For adopted index persons, adoptive parents are registered (Statistics Sweden 2011b).

### **Income, education, social assistance and employment**

Information about: (1) individualized disposable income, (2) completed tertiary education, (3) taxable earned income, (4) social assistance and (5) employment is acquired through Statistics Sweden's longitudinal integration database for health insurance and labour market studies (LISA). The LISA database consists of annual registers from 1990 onwards including all individuals that are 16 years or older. The LISA database is a combination of data from the labour market, educational and social sectors (Statistics Sweden 2011a).

### *4.3 Norway*

All data from Norway used in this project comes from different administrative registers and is made available by Statistics Norway. Statistics Norway collects data from numerous administrative registers for the purposes of producing statistics and making them available to researchers. Data from different registers are linked using a unique personal identifier that is specific to each research project. For more information on the different types of data provided by Statistics Norway, see Statistics Norway (2014).

### **Geocoded data**

Statistics Norway uses information from the Norwegian official address register to identify the place of residence of people registered as residents in Norway. The location of each residence is coded on a 100x100 meter grid covering the entire country. The data on place of residence can be linked to information from other registers by way of a unique personal identifier (Strand and Bloch 2009).

### **Foreign-born**

The information on immigrant status is taken from the National Population Register, made available by Statistics Norway. This register covers everyone who is registered as a resident in

Norway. Foreign-born is here defined as being born abroad to two parents who are born abroad. This definition excludes people born abroad to one or more Norwegian-born parents, but it is in accordance with the standard classification of immigrants used by Statistics Norway. Likewise, descendants are defined as people born in Norway to two foreign-born parents. Immigrants from EU/EFTA are defined as immigrants (as above) born in an EU/EFTA country.

## **Education**

Information on education is taken from the National Education Database, which draws information from several sources, including registers and census data, and is provided by Statistics Norway. Analyses conducted on Norwegian data are based on information on each individual's highest attained educational level on October 1 in a given year. This information is used to construct variables indicating whether or not an individual has completed tertiary education.

## **Income**

Income information is provided by Statistics Norway. Income information is collected from several sources, primarily from registers administered by the Norwegian Tax Administration, and is based on a full sample of the Norwegian population. Depending on the specific indicator, we use information on different income measures, as described in detail below.

## **Social assistance**

The information on social assistance is taken from the income data described above. Here we use information on whether or not an individual received any payments in the form of social assistance or social loans from municipal social services in a given year (Statistics Norway 2017g).

## **Employment**

Employment is defined as having a work income in a given year greater than or equal to two basic amounts (see below). Information on work income is taken from registers maintained by the Norwegian Tax Administration and provided by Statistics Norway. The basic amount is a standard unit in the Norwegian welfare and pension systems, and it is adjusted for inflation each

year. The basic amounts were 56861 NOK in 2003, 66812 NOK in 2007 and 79216 NOK in 2011. This definition excludes some people working part-time at relatively low wages and/ or people with short employment spells. However, this definition allows us to use data on the entire population, covering the entire year.

#### *4.4 Netherlands*

All indicators for the Netherlands have been collected by Statistics Netherlands (CBS). The Social Statistical Database (SSD) is a system of linked statistical registers and surveys which cover a broad range of demographic and socio-economic subjects: from labour force participation to social security, from health care to crime, from housing to migration. The SSD includes detailed information at the personal level. Data are collected from various sources, including population registers of municipalities, tax offices, labour offices and public education institutes. Datasets can be linked to each other through unique personalized identification numbers since 1996. Before that date, data collection was scattered and not standardized. The SSD has expanded widely since 1996. Therefore, some data are available from 1995 onwards, whereas other data are available only since more recent years (Bakker et al. 2014). Detailed information, albeit only in Dutch, on the SSD and its specific sub-datasets are freely available through <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata>.

#### **Geocoded data**

GBA ADRESBUS: for each registered person, all the addresses where one has lived are available, since the start of SSD in 1995. It is known when a person moved to a certain address and when he or she left this address. The database also includes the x and y coordinates for each address. For ResSegr, we have taken the situation at the end of December as the address for that year. Data are derived from the municipal population registers (Statistics Netherlands 2014).

#### **Foreign born**

For constructing the ResSegr indicators, a number of sub-datasets of the SSD have been used. The GBA ('Gemeentelijke Basisadministratie' or Basic Municipal Population Administration)

PERSOONTAB dataset includes personal information on everyone registered in the municipal population registers, including age, country of birth, country of birth of both parents and other personal aspects that do not change over time. This data come from the population registers of municipalities and have been available since the start of SSD (1995) (Statistics Netherlands 2017b).

### **Education**

HOOGSTEOPLTAB: a dataset including a broad collection of data on the highest level of education that a registered person has followed and/or finished. HOOGSTEOPLTAB is made up of data from various education registers and a series of volumes of the Survey on the Working Population (EBB, available since 1996). The dataset has been part of SSD since 1999, but since then, Statistics Netherlands gained access to more and more registers of enrollment and certificates by publicly funded educational institutes. Data are measured each year on the final Friday of September. These data are considered representative for the period until May in the following year. A rise in the level of education is most likely to occur in the period between June and September (the end of the study year) (Statistics Netherlands 2016a).

### **Income**

Integral Personal Income (IPI) database, which is available since 2003. This dataset includes information on personal incomes for all persons registered in The Netherlands, for whom a source of income is known by the national tax office. Data are gathered from national tax registers. The IPI database (part of SSD) includes data on the ‘Socioeconomic category’ of each person in a certain month (‘SECM’). These data are gathered from tax registers and have been available since 2003. The database includes all persons registered in The Netherlands for whom a source of income is known by the national tax office (Statistics Netherlands 2016b).

### **Social assistance and employment**

For social assistance and employment, the sub-dataset ‘SECMBUS’ is used, which describes each individual’s socioeconomic position in a specific month. These socioeconomic positions include being active as a student or an employee, being self-employed, being dependent on

different types of welfare benefits and being otherwise active or inactive. The dataset is longitudinal and ranges from 1999 to 2015 (Statistics Netherlands 2017a).

#### *4.5 Belgium*

##### **Foreign-born**

Indicators of migrants and foreigners in the ResSegr project take into account individuals' country of birth, as well as the country of birth of their parents. For the Belgian dataset, the construction of these indicators draws on two complementary variables: country of birth and family relationship.

The place of birth of every individual with legal residence in Belgium is known from the Census 2011, which was the first census in the country entirely based on administrative sources (see Statistics Belgium 2014). In this case, the place of birth is provided by the National Register of Natural Persons (Registre national des personnes physiques/ Rijksregister van de natuurlijke personen), managed by the Home Affairs Federal Public Service (Service Public Fédéral Intérieur/ Federale Overheidsdienst Binnenlandse Zaken). The Census also includes the refugees in the Waiting Register (Registre d'attente / Wachtregister). The situation on 1 January 2011 is calculated based on the National Register data as on March 1, 2011, in order to control for delays in declaration in the municipalities' records. Moreover, the country of birth is determined according to the international borders applicable on January 1, 2011.

For the construction of indicators 2a to 2f, the Census variable 'place of birth' was complemented with family relationships contained in the National Register. For most Belgian citizens, the National Register provides the country of birth of the mother and the father. For migrants, this information is not always available (see chapter 6 of this report).

##### **Education**

The 'education attainment' is available in the Census 2011 for each individual in Belgium aged 15 or older. It refers to the highest level of education successfully completed by the individual (Statistics Belgium 2014).

The 'education attainment' is the only variable in the Census 2011 that is not entirely based on administrative sources. The primary source for this variable is the information collected in

Census 2001. For the following years, the data were updated with the communities' databases on education created for this purpose. Each linguistic community (French-, Dutch- and German-speaking) is responsible for the education system in their language, and each keep a separate register of the diplomas delivered by the institutions under their responsibility.

For indicators 3a and 3b, the data from the Census 2011 were used in order to identify the individuals who had successfully completed the tertiary education (university diploma).

## **Income**

Income data in Belgium are based on tax records. They are obtained from the IPCAL database (*déclaration d'impôts des personnes physiques / aangifte van de personenbelasting*), managed by the Federal Public Service Finance (Service Public Fédéral Finances / Federale Overheidsdienst Financiën). The individual yearly income available in this database refers to the taxable declared income only, i.e. the sum of income from work (wages or independent businesses), unemployment grants and pension (retirement). It does not include non-taxable revenues (social assistance, child benefits, study grants, etc., nor most income from capital).

## **Social assistance**

Individuals who received social assistance in Belgium in 2011 are those who benefited from the 'minimal subsistence income' (revenue d'integration / leefloon) during the last month of 2010. The minimal subsistence income is granted by the Public Centre for Social Welfare (Centre public d'action sociale/ Openbaar centrum voor maatschappelijk welzijn) to individuals whose resources are not enough for their subsistence and who cannot improve their situation by themselves. Each municipality in Belgium disposes of a Public Centre for Social Welfare. The data concerning the registered individuals across the country are centralised in Crossroad Bank of Social Security (Banque carrefour de la Sécurité sociale/ Kruispuntbank Sociale Zekerheid).

## **Employment**

The variable 'current activity status' from the Census 2011 allows the identification of employed persons in 2011. The category 'employed' in this variable refers to persons aged 15 or older who, during the reference week (the last week of 2010), 'performed at least one hour of work for pay of profit, in cash or in kind', or 'were temporarily absent from a job which they

already worked and to which they maintained a formal attachment, or from a self-employment activity' (Eurostat 2014).

Information on individuals' activity status in the Census 2011 is provided by the Crossroad Bank of Social Security (Banque carrefour de la Sécurité sociale/ Kruispuntbank Sociale Zekerheid), which gathers socioeconomic and labour-related data from several institutions in Belgium (CIN, FAT, INAMI, INASTI, FAMIFED, ONEM, ONP, ONSS, ONSSAPL, SdPSP, Sigedis, SPF Sécurité sociale, SPP Intégration sociale et VDAB/FOREM/ Actiris/ ADG).

## 5. The geographical information

### 5.1 Denmark

The accuracy of the X,Y-coordinates is considered to be of high quality. Most residential addresses in Denmark match one of the address coordinates from OIS (the Public information Server). Of all residential addresses in 2003 and 2011, 0.29 and 0.25 % respectively, of the resident Danish population, could not be distributed geographically at the ratio of 100 x 100 m-square net.

The Danish individual data have been summed together at the level of entry address. Subsequently, each of these addresses has been allocated with a X,Y-coordinate. The addresses have been summed together in 100 x 100 metre square cells by means of the coordinates. The cells have been named in accordance with the coordinates in the lower left-hand corner of the cell – for example, the lower left-hand corner of the square cell “100m\_61691\_7247” has the coordinate values 724700 (X) and 6169100 (Y). The Danish Geodata Agency is the authority which has the general responsibility for the address coordinates. The task of collecting the coordinates for the entry addresses rests with the Danish municipalities. The coordinates are gathered in a joint database containing information on addresses (the public information server (OIS) for property). The Danish address coordinates are calculated in the projection UTM32N ETRS89. The accuracy of the X,Y-coordinates is considered to be of high quality.

### 5.2 Sweden

The basis for the coordinates used in ResSegr is the two-dimensional coordinate system RT 90 2.5 gon V (ESPG:3021). Registered place of residence is derived from the population register at the Swedish Tax Agency and is regulated by Swedish law (Folkbokföringslagen 1991:481, paragraph 6). The registers are considered to be of high quality (The Swedish Tax Agency 2006). For each individual, the north-south and east-west coordinates indicate the centroid of the grid cells where the property where one lives is registered, on December 31 of each year. Individuals that are not registered on a property will therefore have missing values on coordinates. There might also be cases where individuals are registered on a property but in practice live somewhere

else. It has been reported that the quality of the registers in this respect is somewhat poorer for recent home leavers (Statistics Sweden 2008) and students (The Swedish Tax Agency 2006). Some complications of using the property centroid, are that even though the centroid of the property is in one grid, the actual house or apartment building can be located in a neighbouring grid, and that some very large properties might cover a number of grids, resulting in that only the grid with the centroid in, has people registered in it.

The coordinates in GEOSTAR are aggregated to 250 meter squares in urban areas (defined as localities consisting of a group of buildings normally not more than 200 metres apart from each other, and with at least 200 inhabitants and where the proportion of holiday homes is less than 50 percent), and 1000 meter squares outside these urban areas, based on the 2010 urban subdivision (Statistics Sweden 2015a). The urban and rural grids are positioned in such a way that 16 urban grids can fit in one rural grid, so that the centroids of the urban and the rural grids do not overlap.

### *5.3 Norway*

Geo-codes from Norway are provided by Statistics Norway. These codes are based on addresses from the official address register. Statistics Norway links each individual's place of residence to a specific grid cell on a 100x100m grid that covers the entire country. (Strand and Bloch 2009). Each individual is associated with a grid cell by way of a unique personal identifier specific to each research project. This personal identifier makes it possible to link individual anonymous information from various registries to his/her place of residence. The grids are defined in UTM33/WGS84 (EUREF89). Each grid cell is identified by the coordinates in the south-west corner.

Analyses from Norway that cover the entire country and use k-values of 25600 or 51200 are based on geographical data aggregated to a 400x400m grid. This is due to limitations on the search range of the software EquiPop. The program is not able to identify neighborhoods at these high k-values in the north-easternmost parts of the country when the data is based on a 100x100m grid because of a limitation on the number of grid cells it searches through in any direction. The long distances in this part of Norway, combined with the relatively low population density means that we can thus only provide data on aggregated grid cells for these k-values.

#### *5.4 Netherlands*

The coordinates for the Netherlands are based on ‘Rijksdriehoekstelsel’ (RD), which is compatible with the European Terrestrial Reference System 1989 (ETRS89). The x- and y-coordinates of an object are gathered from the Basis administration Addresses and Buildings [Basisregistratie Adressen en Gebouwen] (BAG) and indicate the exact location of a certain address on a map. An object coordinate is made up of a combination of an x-coordinate and a y-coordinate. X-coordinates indicate the east/west location. X-coordinates within The Netherlands are located in-between the values 0 and 300,000. Y-coordinates indicate the north/south location. Y-coordinates within The Netherlands are located in-between the values 289,000 and 629,000. These values are only available at Statistics Netherlands for the Dutch projection area, consisting of the State of The Netherlands and a small area surrounding it in the North Sea, Germany and Belgium. In this component, only objects inside The Netherlands are given a coordinate. The values therefore are restricted to locations on the land and house boat locations on water very close to the land within The Netherlands. The centroid of the RD is a church in the city centre of Amersfoort (Statistics Netherlands 2016c; 2017c).

The 100 x 100 meter grids to be used for EquiPop are named after the coordinates of the bottom left corner.

#### *5.5 Belgium*

Coordinates for Belgium are based on Belgian Lambert 1972 (BD72 system), which is a conformal conic projection constructed on a Hayford ellipsoid (International 1924, with origin at 49°18'0"N, 2°18'0"E). The projection accuracy is considered to be high, with maximum error of 10 cm/km, making it well-suited for a small country such as Belgium.

In the project, the Belgian territory was divided into grids of 100 x 100 metres. The grids' location is defined by the combination of x- and y-coordinates corresponding to their centroids.

The residents inside the grids were identified using the linkage between the National Register population data and the land registry of the General Administration of Patrimonial Documentation managed by the Federal Public Service Finance (Administration Générale de la Documentation Patrimoniale — Cadastre / Algemene Administratie Patrimoniumdocumentatie — Kadaster) in 2011. The latter contains the detailed coordinates of the land parcels and the

buildings within. If a parcel extends upon two grids, their residents are imputed to the grid where the parcel's main building is located.

Approximately 3% of households had missing coordinates in the linkage, although their statistical ward of residence was known. Statistics Belgium operated an 'artificial linkage' for these cases. This uses a probabilistic model in order to attribute an empty dwelling in the same statistical ward to the missing households.

## **6. Detailed description of variables**

### *6.1 Indicator 1a. Share of persons aged 25 and above who have a personal disposable income below 60 % of the median level*

#### **Denmark – Indicator 1a. Share of persons aged 25 and above who have a personal disposable income below 60 % of the median level**

The information about age is based on Statistics Denmark's register of the total population. The register of the total population is daily updated by events including internal and international migration, births and deaths from the Central Population Register (CPR) which originally is registered by the local registers in the 98 municipalities in Denmark. There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the authorities. Analysis has shown that the total population is slightly overestimated (Statistics Denmark (2017a)).

The income data is based on administrative registers from the tax authorities. Disposable income includes income from employment, surplus from own business, interests, dividends, profit from capital investments, transfers and pensions. Obligatory pension schemes, including labour market pensions, are included when paid out, private pension savings are included when earned. From these incomes direct taxes, labour market contributions, interest expenses and alimonies paid are subtracted to form the disposable income. The register has full population coverage.

The disposable income reported for Urban Europe does not include imputed rent and interest paid on mortgages.

#### **Sweden - Indicator 1a. Share of persons aged 25 and above who have a personal disposable income below 60 % of the median level**

The information about age is based on Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the Swedish Tax Agency population register (Folkbokföringregistret). There is some overcoverage in

the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

The individualized disposable income is obtained from Statistics Sweden's Longitudinal integration database for health insurance and labour market studies (LISA). The individualized disposable income variable in LISA is named DispInkPersF for the year 2003 and DispInkPersF04 for the year 2011. The difference between DispInkPersF and DispInkPersF04 is that the consumption weights that are used for the calculation were changed for the 2011 variable. DispInkPersF04 and DispInkPersF variables are obtained by dividing the sum of all the family members' individual disposable incomes, and then multiplied by the individuals' consumption weights. Negative values have been set to zero. DispInkPersF04 and DispInkPersF are registered on the December 31 for each year (Statistics Sweden 2011a).

The median value that is used in the definition of the variable 1a for this project is based on the individualized disposable income for those who are 25 years old or older.

#### **Norway - Indicator 1a. Share of persons aged 25 and above who have a personal disposable income below 60 % of the median level**

Age is calculated from birth year, which is available from the population registry data provided by Statistics Norway. Note that we use the age on January 1 each year.

Personal disposable income is based on tax return data provided by Statistics Norway. Personal disposable income is measured by the variable "wies", defined as the sum of occupational income, capital income, taxable transfers and tax-free transfers, minus taxes and negative transfers (Statistics Norway 2017f).

The median personal disposable income is calculated from the personal disposable income distribution of people in the age group 25 and above who are registered as resident that year.

#### **The Netherlands - Indicator 1a. Share of persons aged 25 and above who have a personal disposable income below 60 % of the median level**

Data on age has been derived from the municipal population registers (GBA PERSOONTAB), which include the year of birth (Statistics Netherlands 2017b).

Data on income are derived from the Integral Personal Income (IPI) database, which includes information on personal incomes for all persons registered in The Netherlands, for whom a

source of income is known by the national tax office. The ‘personal income’ variable used for the indicators consists of personal gross wages, minus paid income insurances. This comes down to the income of an individual before tax that can be attributed solely to this person, and thus excludes subsidies that can be attributed to the entire household. The limit of 60% of the median level was at 14220 in 2003 and 17340 in 2011.

Since the data are derived from tax registers, the quality and reliability is in in general very high. The data are integral for all persons that are registered in The Netherlands, which ensures a representative research population.

However, the database does not include all sources of income. Sources that cannot be linked to individual household members, such as rent subsidies and income from capital, are not included in the personal income.

Personal income is not equal to welfare. A single person with only AOW (pensions) earns 70 percent of the amount that couples would earn together. In contrast, standardized household income is equal for both households.

Furthermore, negative incomes are possible (0.5 percent of the population). This is often related to surpluses of self-employed persons or with incomes from other sources.

There are no breaks in the data; the IPI databases (part of SSD) is measured the same way in both 2003 and 2011 (Statistics Netherlands 2016b).

### **Belgium - Indicator 1a. Share of persons aged 25 and above who have a personal disposable income below 60 % of the median level**

Income data in Belgium, based on tax records, is expected to cover all individuals’ declared (‘legal’) economic earnings. The data quality is considered to be good, as the Federal Public Service Finance proceed to consistency checks of tax returns.

The dataset covers taxable income only, that is, it does not include non-taxable earnings (social assistance, child benefits, study grants, etc., as well as most income from capital).

Indicator 1a should be based on disposable income after social transfer. However, the distinction between taxable net earnings and disposable income could not be made in the Belgian dataset for technical reasons. As a consequence, indicator 1a does not include all social transfers (such as social assistance and child benefits). We expect this to have little impact on the quality of the indicators because they rely on relative measures of income — the median level (1a).

## *6.2 Indicator 2a. Immigrants' share of the population*

### **Denmark - Indicator 2a. Immigrants' share of the population**

An immigrant is defined as a person who is born abroad and whose parents are both (or one in case we only have information of the either the father or mother) foreign citizens or born abroad. If no information available about the father or mother the person is born abroad he/ she is taken as an immigrant. Immigrants with parents born in Denmark who have foreign citizenship are included. Calculations by Statistics Denmark show that the number of migrants according to this definition is 0.7 % higher compared to the project definition. The quality of the classification of persons as immigrants is generally high. The classification is based on the registration of country of origin and citizenship in the Central Population Register (CPR).

Country of birth is obtained from the CPR run by the ministry of Social and Interior, on January 1 every year. The 98 municipalities register country of birth for all people who immigrate and take residence in the municipality. Municipalities use the register for administrative purposes and have an interest in keeping the quality as high as possible.

Information about the parents in order to define if the person is an immigrant or descendent is available for Statistics Denmark in the population register which source is the CPR. CPR links parents with their (adopted) children.

There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the authorities. Analysis has shown that the population is slightly overestimated (Statistics Denmark (2017a)). For immigrants, analyses show that the overcoverage is close to 1 %.

### **Sweden - Indicator 2a. Immigrants' share of the population**

Immigrants are defined in this project as: someone who is not born in Sweden and who does not have a parent born in Sweden.

Country of birth is obtained from Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the Swedish Tax Agency population register (Folkbokföringregistret). There is some overcoverage in

the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

Information about who are a person's parents is collected through the Swedish multi-generation register (Flergenerationsregistret). The multi-generation register is based on the Swedish Tax Agency population register and is a register of persons who have been registered in Sweden at some time since 1961 and who were born in 1932 or later – so-called index persons. The register contains connections between index persons and their biological parents. For adopted index persons, adoptive parents are registered. The coverage for index persons is almost complete from 1968 onwards. For the period 1961 to 1967, coverage is good, but not as comprehensive (Statistics Sweden 2011b).

If no information exists on the birthplace of the parents, the person is considered an immigrant if he/she is foreign-born.

#### **Norway - Indicator 2a. Immigrants' share of the population**

Data on immigrant background is taken from the population register, and provided by Statistics Norway. This register contains information on each individual's immigrant category. Immigrants are defined as people born abroad (meaning that the mother was not registered as resident in Norway at the time of birth), with two foreign-born parents. This definition excludes people born abroad with one or more Norwegian-born parents. The data covers everyone registered as resident in Norway. This means that people who reside in Norway without being registered as settled/ resident (due to short stays, illegal immigration, delays in registration, because they await a decision on an application for asylum, or for other reasons) are not included. Further, people who emigrate without informing the authorities may not be excluded from the registers (Statistics Norway 2017b).

The specific variable used to identify immigrants in the data is called "invkat", and immigrants are identified by the value "B".

#### **The Netherlands - Indicator 2a. Immigrants' share of the population**

The indicators on migrants are based on data from the Dutch population register regarding the country of birth of a person and (if known) both parents' country of birth (GBA PERSOONTAB).

- Country of birth: the country where a person was born, as mentioned in the municipal population registers.
- Country of birth father: the country where a person's father was born, as mentioned in the municipal population registers
- Country of birth mother: country where a person's mother was born, as mentioned in the municipal population registers

Each country in the world has a code. Code 6030 indicates a person is born in The Netherlands. All other codes indicate that a person is born abroad. The definition of a migrant for this indicator is therefore as follows:

- A person is classified as an immigrant if the code for the country of birth is not equal to '6030' (The Netherlands), the code for country of birth of the father is not equal to '6030' and the code for the country of birth of the mother is not equal to '6030'.
- A person is not classified as an immigrant if the code for the country of birth is equal to '6030', or the code for the country of birth is equal to '6030' but the code for the country of birth of the father and/or the mother is equal to '6030'.

The quality of the data from the municipal registers is very high. The data are integral and cover all persons that were or have been included in the municipal registers since 1995. This guarantees a research population that is representative for the Dutch population.

The data are fully comparable from 1995, the first year the statistics were released, onward. There are thus no breaks between the 2003 and 2011 data files (Statistics Netherlands 2017b).

### **Belgium - Indicator 2a. Immigrants' share of the population**

Indicator 2a includes all persons not born in Belgium whose parents are also foreign-born. Its computation used individuals' country of birth and kinship. The country of birth of every individual with legal residence in Belgium is known from the 2011 Census variable 'place of birth'. In order to control for the country of births of the parents, this variable was complemented with family relationships contained in the National Register.

There are two potential sources of bias in the construction of this indicator for Belgium.

The first one relates to the coverage of the National Register of Natural Persons, from which the population in 2011 is drawn. On the one hand, there is under-declaration of departures, that is, Belgians and foreigners who leave the country and do not report their out-migration to the

municipal administration. On the other hand, there is under-declaration of arrivals, both from irregular immigrants, and from European workers living in Belgium who choose not to establish their main residence in the country. However, these coverage issues are expected to have little impact on the population data, since municipalities proceed to regular verification of their records.

The second potential source of bias concerns the parents' origin. In fact, the country of birth of the parents is unknown for many foreigners. If there is no information on parent's origin, a foreign born was considered an immigrant. However, in this case, missing values of the parents' country of birth should not bias the results: if one of the parents was born in Belgium, their country of birth is likely to be known because they are directly linked in the National Register of Natural Persons.

### *6.3 Indicator 2b. EU28/EFTA Immigrants' share of the population*

#### **Denmark - Indicator 2b. EU28/EFTA Immigrants' share of the population**

An immigrant is defined as a person who is born abroad and whose parents are both (or one in case we only have information of the either the father or mother) foreign citizens or born abroad. If no information available about the father or mother the person is born abroad he/ she is taken as an immigrant. Immigrants with parents born in Denmark who have foreign citizenship are included. Calculations by Statistics Denmark show that the number of migrants according to this definition is 0.7 % higher compared to the project definition.

The quality of the classification of persons as immigrants is generally high. The classification is based on the registration of country of origin and citizenship in the Central Population Register (CPR).

Country of birth is obtained from the CPR run by the ministry of Social and Interior, on January 1 every year. The 98 municipalities register country of birth for all people who immigrate and take residence in the municipality. Municipalities use the register for administrative purposes and have an interest in keeping the quality as high as possible.

Information about the parents in order to define if the person is an immigrant or descendant is available for Statistics Denmark in the population register which source is the CPR. CPR links parents with their (adopted) children.

There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the authorities. Analysis has shown that the population is slightly overestimated (Statistics Denmark (2017a)). For immigrants, analyses show that the overcoverage is close to 1 %.

Information for each EU/EFTA country is available for all immigrants. The delimitation of the EU/EFTA group is rather precise.

### **Sweden - Indicator 2b. EU28/EFTA Immigrants' share of the population**

The Swedish project team only had access to grouped information on country of birth. This meant that the definition of an EU28/EFTA immigrant for Sweden is defined as: someone who is born in the EU28 region or in one of the Nordic countries except for Sweden, and who does not have a parent born in Sweden. The Swedish definition thus lacks immigrants from Switzerland and Lichtenstein in this variable. This should not have any large effects on the variable as the number of people registered in Sweden that is born in Switzerland and Lichtenstein in the years 2003 and 2011 is relatively small, see table 1.

Table 1. Number of persons registered in Sweden born in Lichtenstein and Switzerland in 2003 and 2011

	2003	2011
Lichtenstein	2	1
Switzerland	2557	3211

Source: Statistics Sweden 2017

Country of birth is obtained from Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the Swedish Tax Agency population register (Folkbokföringregistret). There is some overcoverage in the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

Information about who are a person's parents is collected through the Swedish multi-generation register (Flergenerationsregistret). The multi-generation register is based on the Swedish Tax Agency population register and is a register of persons who have been registered in Sweden at some time since 1961 and who were born in 1932 or later – so-called index persons. The register contains connections between index persons and their biological parents. For adopted index persons, adoptive parents are registered. The coverage for index persons is almost complete from 1968 onwards. For the period 1961 to 1967, coverage is good, but not as comprehensive (Statistics Sweden 2011b).

If no information exists on the birthplace of the parents the person is considered an immigrant if he/she is foreign-born.

### **Norway - Indicator 2b. EU28/EFTA Immigrants' share of the population**

The definition of immigrants here is identical to the definition described under indicator 1a, with the additional restriction that the mother's country of residence at birth is a EU28/EFTA country (Statistics Norway 2017b). Immigrants are defined as people born abroad, with two foreign-born parents. This definition excludes people born abroad with one or more Norwegian-born parents.

The specific variable used to identify immigrants in the data is called "invkat", and immigrants are identified by the value "B". Information on the mother's country of residence at birth is given by the variable "fodeland". EU28/EFTA countries are defined by the values 000 (Norway), 101 (Denmark), 103 (Finland), 105 (Iceland), 106 (Sweden), 112 (Belgium), 113 (Bulgaria), 115 (Estonia), 117 (France), 119 (Greece), 121 (Ireland), 122 (Croatia), 123 (Italy), 124 (Latvia), 126 (Malta), 127 (Netherlands), 128 (Liechtenstein), 129 (Luxembourg), 131 (Poland), 132 (Portugal), 133 (Romania), 136 (Lithuania), 137 (Spain), 139 (United Kingdom), 141 (Switzerland), 144 (Germany), 146 (Slovenia), 152 (Hungary), 153 (Austria), 157 (Slovakia), 158 (Czech Republic) and 500 (Cyprus). Note that this means that 102 (Greenland), 104 (Faroe Islands), 114 (Andorra), 118 (Gibraltar), 130 (Monaco), 134 (San Marino) and 154 (Vatican City State) are not included as EU28/EFTA countries.

The data covers everyone registered as a resident in Norway. This means that people who reside in Norway without being registered as settled (due to short stays, illegal immigration, delays in registration, because they await a decision on an application for asylum, or for other

reasons) are not counted. Further, people who emigrate without informing the authorities may not be excluded from the registers (Statistics Norway 2017b).

### **The Netherlands - Indicator 2b. EU28/EFTA Immigrants' share of the population**

The indicators on migrants are based on data from the Dutch population register regarding the country of birth of a person and (if known) both parents' country of birth (GBA PERSOONTAB).

- Country of birth: the country where a person was born, as mentioned in the municipal population registers.
- Country of birth father: the country where a person's father was born, as mentioned in the municipal population registers
- Country of birth mother: country where a person's mother was born, as mentioned in the municipal population registers

Each country in the world has a code. Code 6030 indicates a person is born in The Netherlands. All other codes indicate that a person is born abroad. The definition of a migrant for this indicator is therefore as follows:

- A person is classified as an immigrant if the code for the country of birth is not equal to '6030' (The Netherlands), the code for country of birth of the father is not equal to '6030' and the code for the country of birth of the mother is not equal to '6030'.
- Based on the country codes, we could make a distinction between EU/EFTA and non-EU/EFTA migrants. If the code for a person's country of birth is associated with one of the 28 EU countries (excluding The Netherlands) or one of the four EFTA countries, this person is categorised as an 'EU/EFTA immigrant'.
- A person is not classified as an immigrant if the code for the country of birth is equal to '6030', or the code for the country of birth is equal to '6030' but the code for the country of birth of the father and/or the mother is equal to '6030'.

The quality of the data from the municipal registers is very high. The data are integral and cover all persons that were or have been included in the municipal registers since 1995. This guarantees a research population that is representative for the Dutch population.

The data are fully comparable from 1995, the first year the statistics were released, onward. There are thus no breaks between the 2003 and 2011 data files (Statistics Netherlands 2017b).

## **Belgium - Indicator 2b. EU28/EFTA Immigrants' share of the population**

Indicator 2b includes all persons born in a EU28/EFTA country and whose parents were not born in Belgium. Its computation used individuals' country of birth and kinship. The country of birth of every individual with legal residence in Belgium is known from the 2011 Census variable 'place of birth'. In order to control for the country of births of the parents, this variable was complemented with family relationships contained in the National Register.

There are two potential sources of bias in the construction of this indicator for Belgium.

The first one relates to the coverage of the National Register of Natural Persons, from which the population in 2011 is drawn. On the one hand, there is under-declaration of departures, that is, Belgians and foreigners who leave the country and do not report their out-migration to the municipal administration. On the other hand, there is under-declaration of arrivals, both from irregular immigrants, and from European workers living in Belgium who choose not to establish their main residence in the country. However, these coverage issues are expected to have little impact on the population data, since municipalities proceed to regular verification of their records.

The second potential source of bias concerns the parents' origin. In fact, the country of birth of the parents is unknown for many foreigners. If there is no information on parent's origin, a foreign born was considered an immigrant. However, in this case, missing values of the parents' country of birth should not bias the results: if one of the parents was born in Belgium, their country of birth is likely to be known because they are directly linked in the National Register of Natural Persons.

### *6.4 Indicator 2c. NON-EU28/EFTA Immigrants' share of the population*

## **Denmark – Indicator 2c. NON-EU/EFTA Immigrants' share of the population**

An immigrant is defined as a person who is born abroad and whose parents are both (or one in case we only have information of the either the father or mother) foreign citizens or born abroad. If no information available about the father or mother the person is born abroad he/ she is taken as an immigrant. Immigrants with parents born in Denmark who have foreign citizenship

are included. Calculations by Statistics Denmark show that the number of migrants according to this definition is 0.7 % higher compared to the project definition. The quality of the classification of persons as immigrants is generally high. The classification is based on the registration of country of origin and citizenship in the Central Population Register (CPR).

Country of birth is obtained from the CPR run by the ministry of Social and Interior, on January 1 every year. The 98 municipalities register country of birth for all people who immigrate and take residence in the municipality. Municipalities use the register for administrative purposes and have an interest in keeping the quality as high as possible.

Information about the parents in order to define if the person is an immigrant or descendant is available for Statistics Denmark in the population register which source is the CPR. CPR links parents with their (adopted) children.

There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the authorities. Analysis has shown that the population is slightly overestimated (Statistics Denmark (2017a)). For immigrants, analyses show that the overcoverage is close to 1 %.

Information for each NON-EU/EFTA country is available for all immigrants. Unknown country is part of the NON-EU/EFTA group. The delimitation of the NON-EU/EFTA group is rather precise.

### **Sweden – Indicator 2c. NON-EU/EFTA Immigrants' share of the population**

The Swedish project team only had access to grouped information on country of birth. This meant that the definition of a non-EU28/EFTA immigrant for Sweden is defined as: someone who is not born in the EU28 region nor in one of the Nordic countries, and who does not have a parent born in Sweden. The Swedish definition thus includes immigrants from Switzerland and Lichtenstein in this variable. This should not have any large effects on the variable as the number of people registered in Sweden that is born in Switzerland and Lichtenstein in the years 2003 and 2011 is relatively small, see table 1.

Country of birth is obtained from Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the Swedish Tax Agency population register (Folkbokföringregistret). There is some overcoverage in

the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

Information about who are a person's parents is collected through the Swedish multi-generation register (Flergenerationsregistret). The multi-generation register is based on the Swedish Tax Agency population register and is a register of persons who have been registered in Sweden at some time since 1961 and who were born in 1932 or later – so-called index persons. The register contains connections between index persons and their biological parents. For adopted index persons, adoptive parents are registered. The coverage for index persons is almost complete from 1968 onwards. For the period 1961 to 1967, coverage is good, but not as comprehensive (Statistics Sweden 2011b).

If no information exists on the birthplace of the parents the person is considered an immigrant if he/she is foreign-born.

### **Norway – Indicator 2c. NON-EU/EFTA Immigrants' share of the population**

Immigrants from non-EU/EFTA-countries are defined similarly to immigrants from EU/EFTA-countries. Specifically, immigrants from non-EU/EFTA-countries are defined as people born abroad (meaning that the mother was not registered as resident in Norway at the time of birth) to two foreign-born parents, and whose mother was not resident in an EU/EFTA-country at the time of birth. The specific variable used to identify immigrants in the data is called "invkat", and immigrants are identified by the value "B". Information on the mother's country of residence at birth is given by the variable "fodeland". EU28/EFTA countries are defined by the values 000 (Norway), 101 (Denmark), 103 (Finland), 105 (Iceland), 106 (Sweden), 112 (Belgium), 113 (Bulgaria), 115 (Estonia), 117 (France), 119 (Greece), 121 (Ireland), 122 (Croatia), 123 (Italy), 124 (Latvia), 126 (Malta), 127 (Netherlands), 128 (Liechtenstein), 129 (Luxembourg), 131 (Poland), 132 (Portugal), 133 (Romania), 136 (Lithuania), 137 (Spain), 139 (United Kingdom), 141 (Switzerland), 144 (Germany), 146 (Slovenia), 152 (Hungary), 153 (Austria), 157 (Slovakia), 158 (Czech Republic) and 500 (Cyprus). Note that this means that 102 (Greenland), 104 (Faroe Islands), 114 (Andorra), 118 (Gibraltar), 130 (Monaco), 134 (San Marino) and 154 (Vatican City State) are not included as EU28/EFTA countries.

The data covers everyone registered as resident in Norway. This means that people who reside in Norway without being registered as settled (due to short stays, illegal immigration,

delays in registration, because they await a decision on an application for asylum, or for other reasons) are not counted. Further, people who emigrate without informing the authorities may not be excluded from the registers (Statistics Norway 2017b).

### **The Netherlands – Indicator 2c. NON-EU/EFTA Immigrants’ share of the population**

The indicators on migrants are based on data from the Dutch population register regarding the country of birth of a person and (if known) both parents’ country of birth (GBA PERSOONTAB).

- Country of birth: the country where a person was born, as mentioned in the municipal population registers.
- Country of birth father: the country where a person's father was born, as mentioned in the municipal population registers
- Country of birth mother: country where a person's mother was born, as mentioned in the municipal population registers

Each country in the world has a code. Code 6030 indicates a person is born in The Netherlands. All other codes indicate that a person is born abroad. The definition of a migrant for this indicator is therefore as follows:

- A person is classified as an immigrant if the code for the country of birth is not equal to ‘6030’ (The Netherlands), the code for country of birth of the father is not equal to ‘6030’ and the code for the country of birth of the mother is not equal to ‘6030’.
- Based on the country codes, we could make a distinction between EU/EFTA and non-EU/EFTA migrants. In case the code for a person’s country of birth is associated with a country other than one of the 28 EU countries (including The Netherlands) or one of the four EFTA countries, this person is categorised as a ‘non-EU/EFTA immigrant’.
- A person is not classified as an immigrant if the code for the country of birth is equal to ‘6030’, or the code for the country of birth is equal to ‘6030’ but the code for the country of birth of the father and/or the mother is equal to ‘6030’.

The quality of the data from the municipal registers is very high. The data are integral and cover all persons that were or have been included in the municipal registers since 1995. This guarantees a research population that is representative for the Dutch population.

The data are fully comparable from 1995, the first year the statistics were released, onward. There are thus no breaks between the 2003 and 2011 data files (Statistics Netherlands 2017b).

### **Belgium – Indicator 2c. NON-EU/EFTA Immigrants’ share of the population**

Indicator 2c includes all persons born in a non-EU28/EFTA country and whose parents were not born in Belgium. The calculation used individuals’ country of birth and kinship. The country of birth of every individual with legal residence in Belgium is known from the 2011 Census variable ‘place of birth’. In order to control for the country of births of the parents, this variable was complemented with family relationships contained in the National Register.

There are two potential sources of bias in the construction of this indicator for Belgium.

The first one relates to the coverage of the National Register of Natural Persons, from which the population in 2011 is drawn. On the one hand, there is under-declaration of departures, that is, Belgians and foreigners who leave the country and do not report their out-migration to the municipal administration. On the other hand, there is under-declaration of arrivals, both from irregular immigrants, and from European workers living in Belgium who choose not to establish their main residence in the country. However, these coverage issues are expected to have little impact on the population data, since municipalities proceed to regular verification of their records.

The second potential source of bias concerns the parents’ origin. In fact, the country of birth of the parents is unknown for many foreigners. If there is no information on parent’s origin, a foreign born was considered an immigrant. However, in this case, missing values of the parents’ country of birth should not bias the results: if one of the parents was born in Belgium, their country of birth is likely to be known because they are directly linked in the National Register of Natural Persons.

*6.5 Indicator 2d. Share of the population for persons with foreign origin (immigrants+descendent)*

**Denmark – Indicator 2d. Share of the population for persons with foreign origin (immigrants+descendent)**

An immigrant is defined as a person who is born abroad and whose parents are both (or one in case we only have information of the either the father or mother) foreign citizens or born abroad. Immigrants with parents born in Denmark but who have foreign citizenship are defined as immigrants. If there is no information available about the parents and the person is born in Denmark but is a foreign citizen, he/ she is taken as a descendent. Calculations by Statistics Denmark show that the number of persons with foreign originimmigrants according to this definition is 0.7 % higher compared to the project definition.

A descendant is defined as a person born in Denmark and whose parents (or one of them if only information about the father or mother is available) have foreign citizenship and are immigrants. If we have no information about the parents and the person is born in Denmark and is a foreign citizen, he/ she is taken as a descendant. Calculation based on recent data show that the indicator include 1.6 % extra persons, compared to the project definition for this indicator.

The quality of the classification of persons as immigrants and descendants is generally high. The classification is based on the registration of country of origin and citizenship in the Central Population Register (CPR).

Country of birth is obtained from the CPR run by the ministry of Social and Interior, on January 1 every year. The 98 municipalities register country of birth for all people who immigrate and take residence in the municipality. Municipalities use the register for administrative purposes and have an interest in keeping the quality as high as possible.

Information about the parents in order to define if the person is an immigrant or descendant is available for Statistics Denmark in the population register which source is the CPR. CPR links parents with their (adopted) children.

There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the authorities. Analysis has shown that the population is slightly overestimated (Statistics Denmark (2017a)). For immigrants and descendants, analyses show that the overcoverage is close to 1 %.

**Sweden – Indicator 2d. Share of the population for persons with foreign origin (immigrants+descendent)**

Foreign origin is defined as someone who is not born in Sweden and who does not have a parent born in Sweden, plus anyone born in Sweden who does not have parents or a parent born in Sweden.

Country of birth is obtained from Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the Swedish Tax Agency population register (Folkbokföringregistret). There is some overcoverage in the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

Information about who are a person's parents is collected through the Swedish multi-generation register (Flergenerationsregistret). The multi-generation register is based on the Swedish Tax Agency population register and is a register of persons who have been registered in Sweden at some time since 1961 and who were born in 1932 or later – so-called index persons. The register contains connections between index persons and their biological parents. For adopted index persons, adoptive parents are registered. The coverage for index persons is almost complete from 1968 onwards. For the period 1961 to 1967, coverage is good, but not as comprehensive (Statistics Sweden 2011b).

If no information exists on the birthplace of the parents the person is considered an immigrant if he/she is foreign-born.

**Norway – Indicator 2d. Share of the population for persons with foreign origin (immigrants+descendent)**

Data on immigrant background is taken from the population register, and provided by Statistics Norway. This register contains information on each individual's immigrant category. Immigrants are defined as people born abroad (meaning that the mother was not registered as resident in Norway at the time of birth), with two foreign-born parents. This definition excludes people born abroad with one or more Norwegian-born parents. Descendants are defined as people born in Norway with two foreign-born parents. This definition excludes people born in Norway with one Norwegian-born parent.

The data covers everyone registered as a resident in Norway. This means that people who reside in Norway without being registered as settled (due to short stays, illegal immigration, delays in registration, because they await a decision on an application for asylum, or for other reasons) are not counted. Further, people who emigrate without informing the authorities may not be excluded from the registers (Statistics Norway 2017b).

The specific variable used to identify immigrants and descendants in the data is called “invkat”, and immigrants are identified by the value “B”, while descendants are identified by the value “C”.

### **The Netherlands – Indicator 2d. Share of the population for persons with foreign origin (immigrants+descendent)**

Information on descendants come from the GBA PERSOONTAB dataset and is based on country of birth.

Statistics Netherlands generally does not use the country of birth (of person and parents) indicators to define ‘category of origin’ in its own publications. Rather, it uses its own subdivision of origin groups and migrant generations, which does not totally overlap with the definitions used in ResSegr.

- Category of origin (Statistics Netherlands definition): the country a person has a connection with, based on the country of birth of either one of the parents or him or herself. Explanation: ‘autochtoon’ (native Dutch): a person of whom both parents were born in The Netherlands. First generation ‘allochtoon’: a person born abroad with at least one parent born abroad. Second generation ‘allochtoon’: a person born in The Netherlands with at least one parent born abroad. An ‘autochtoon’ has The Netherlands as category of origin. For a first generation ‘allochtoon’, the country of birth is his or her category of origin. For a second generation ‘allochtoon’ the category of origin is the mother’s country of birth, unless this is The Netherlands as well. In that case the category of origin is determined by the father’s country of birth.

The most important difference is that Statistics Netherlands considers persons with one Dutch and one foreign parent as descendants, whereas these persons are seen as natives in the ResSegr definition. Therefore the countries of birth of person and parents are used to construct the indicators for migrants and foreign origin according to the ResSegr definition.

The definition of a migrant for this indicator is therefore as follows:

- A person is classified as an immigrant if the code for the country of birth is not equal to '6030' (The Netherlands), the code for country of birth of the father is not equal to '6030' and the code for the country of birth of the mother is not equal to '6030'.
- A person is not classified as an immigrant if the code for the country of birth is equal to '6030', or the code for the country of birth is equal to '6030' but the code for the country of birth of the father and/or the mother is equal to '6030'.

The definition for a descendant for this indicator is as follows:

- A person is classified as a descendant if the code for country of birth is equal to '6030' (The Netherlands) but the country of birth of both parents is not equal to '6030'.

A person is considered of foreign origin if he or she can be assigned to either immigrants or descendants, according to the abovementioned definitions.

The quality of the data from the municipal registers is very high. The data are integral and cover all persons that were or have been included in the municipal registers since 1995. This guarantees a research population that is representative for the Dutch population.

The data are fully comparable from 1995, the first year the statistics were released, onward. There are thus no breaks between the 2003 and 2011 data files (Statistics Netherlands 2017b).

### **Belgium – Indicator 2d. Share of the population for persons with foreign origin (immigrants+descendent)**

Indicator 2d includes all immigrants: persons not born in Belgium and whose parents were also not born in Belgium, and all descendants: persons born in Belgium whose mother is foreign-born (or father, if information on the mother is missing). The calculation used individuals' country of birth and kinship. The country of birth of every individual with legal residence in Belgium is known from the 2011 Census variable 'place of birth'. In order to control for the country of births of the parents, this variable was complemented with family relationships contained in the National Register.

There are two potential sources of bias in the construction of the indicators of immigrants and persons with foreign origin for Belgium.

The first one relates to the coverage of the National Register of Natural Persons, from which the population in 2011 is drawn. On the one hand, there is under-declaration of departures, that is,

Belgians and foreigners who leave the country and do not report their out-migration to the municipal administration. On the other hand, there is under-declaration of arrivals, both from irregular immigrants, and from European workers living in Belgium who choose not to establish their main residence in the country. However, these coverage issues are expected to have little impact on the population data, since municipalities proceed to regular verification of their records.

The second potential source of bias concerns the parents' origin. If there is no information on the parents' country of birth, a native-born person is not considered as a descendant. As 32% of Belgian-born individuals have both parents of unknown origin, these were not included among the descendants, making up for an underestimation of foreign-origin persons in indicator 2d. Statistics Belgium is currently working on an algorithm to identify parents' origin; once the new data are ready we expect to update this indicator and obtain more accurate measures.

*6.6 Indicator 2e. Share of the population for persons with foreign origin (immigrants+descendent) in an EU28/EFTA country*

**Denmark – Indicator 2e. Share of the population for persons with foreign origin (immigrants+descendants) in an EU28/EFTA country**

An immigrant is defined as a person who is born abroad and whose parents are both (or one in case we only have information of the either the father or mother) foreign citizens or born abroad. Immigrants with parents born in Denmark but who have foreign citizenship are defined as immigrants. If there is no information available about the parents and the person is born in Denmark but is a foreign citizen, he/ she is taken as a descendent. Calculations by Statistics Denmark show that the number of persons with foreign originimmigrants according to this definition is 0.7 % higher compared to the project definition.

A descendant is defined as a person born in Denmark and whose parents (or one of them if only information about the father or mother is available) have foreign citizenship and are immigrants. If we have no information about the parents and the person is born in Denmark and is a foreign citizen, he/ she is taken as a descendant. Calculation based on recent data show that the indicator include 1.6 % extra persons, compared to the project definition for this indicator.

The quality of the classification of persons as immigrants and descendants is generally high. The classification is based on the registration of country of origin and citizenship in the Central Population Register (CPR).

Country of birth is obtained from the CPR run by the ministry of Social and Interior, on January 1 every year. The 98 municipalities register country of birth for all people who immigrate and take residence in the municipality. Municipalities use the register for administrative purposes and have an interest in keeping the quality as high as possible.

Information about the parents in order to define if the person is an immigrant or descendant is available for Statistics Denmark in the population register which source is the CPR. CPR links parents with their (adopted) children.

There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the authorities. Analysis has shown that the population is slightly overestimated (Statistics Denmark (2017a)). For immigrants and descendants, analyses show that the overcoverage is close to 1 %.

Information for each EU/EFTA country is available for all immigrants and descendants. The delimitation of the EU/EFTA group is rather precise.

### **Sweden – Indicator 2e. Share of the population for persons with foreign origin (immigrants+descendants) in an EU28/EFTA country**

The Swedish project team only had access to grouped information on country of birth. This meant that a person with foreign origin (immigrants+descendants) in a EU28/EFTA country is defined as: someone who is born in the EU28 region or one of the Nordic countries except for Sweden and who does not have a parent born in Sweden, plus those born in Sweden without a parent born in Sweden, but with parents or a parent born in the EU28 region or one of the Nordic countries, except for Sweden.

Because of this, Sweden's definition of the 2e variables does not include immigrants with foreign origin in Switzerland and Lichtenstein. This however should not have any large effects as the number of people registered in Sweden that is born in Switzerland and Lichtenstein in the years 2003 and 2011 is relatively small, see table 1.

Country of birth is obtained from Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the

Swedish Tax Agency population register (Folkbokföringregistret). There is some overcoverage in the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

Information about who are a person's parents is collected through the Swedish multi-generation register (Flergenerationsregistret). The multi-generation register is based on the Swedish Tax Agency population register and is a register of persons who have been registered in Sweden at some time since 1961 and who were born in 1932 or later – so-called index persons. The register contains connections between index persons and their biological parents. For adopted index persons, adoptive parents are registered. The coverage for index persons is almost complete from 1968 onwards. For the period 1961 to 1967, coverage is good, but not as comprehensive (Statistics Sweden 2011b).

If no information exists on the birthplace of the parents, the person is considered an immigrant if he/ she is foreign-born.

#### **Norway – Indicator 2e. Share of the population for persons with foreign origin (immigrants+descendants) in an EU28/EFTA country**

Data on immigrant background is taken from the population register, and provided by Statistics Norway. This register contains information on each individual's immigrant category. Immigrants are defined as people born abroad (meaning that the mother was not registered as resident in Norway at the time of birth), with two foreign-born parents. This definition excludes people born abroad with one or more Norwegian-born parents. Descendants are defined as people born in Norway with two foreign-born parents. This definition excludes people born in Norway with one Norwegian-born parent.

The specific variable used to identify immigrants and descendants in the data is called "invkat", and immigrants are identified by the value "B", while descendants are identified by the value "C".

Immigrants from EU/EFTA-countries are defined as immigrants whose mother was resident in an EU/EFTA-country at the time of birth. Descendants of immigrants from from EU/EFTA-countries are defined as descendants whose parents were born in an EU/EFTA-country. In cases where the parents have different countries of birth, origin is defined by the mother's country of birth (Statistics Norway 2017c).

For immigrants, information on the mother's country of residence at birth is given by the variable "fodeland". For descendants, information on parents' country of birth is given by the variable "landbak3gen". Both of these variables follow the same coding. EU28/EFTA countries are defined by the values 000 (Norway), 101 (Denmark), 103 (Finland), 105 (Iceland), 106 (Sweden), 112 (Belgium), 113 (Bulgaria), 115 (Estonia), 117 (France), 119 (Greece), 121 (Ireland), 122 (Croatia), 123 (Italy), 124 (Latvia), 126 (Malta), 127 (Netherlands), 128 (Liechtenstein), 129 (Luxembourg), 131 (Poland), 132 (Portugal), 133 (Romania), 136 (Lithuania), 137 (Spain), 139 (United Kingdom), 141 (Switzerland), 144 (Germany), 146 (Slovenia), 152 (Hungary), 153 (Austria), 157 (Slovakia), 158 (Czech Republic) and 500 (Cyprus). Note that this means that 102 (Greenland), 104 (Faroe Islands), 114 (Andorra), 118 (Gibraltar), 130 (Monaco), 134 (San Marino) and 154 (Vatican City State) are not included as EU28/EFTA countries.

The data covers everyone registered as resident in Norway. This means that people who reside in Norway without being registered as settled (due to short stays, illegal immigration, delays in registration, because they await a decision on an application for asylum, or for other reasons) are not counted. Further, people who emigrate without informing the authorities may not be excluded from the registers (Statistics Norway 2017b).

### **The Netherlands – Indicator 2e. Share of the population for persons with foreign origin (immigrants+descendants) in an EU28/EFTA country**

Information on descendants come from the GBA PERSOONTAB dataset and is based on country of birth.

Statistics Netherlands generally does not use the country of birth (of person and parents) indicators to define 'category of origin' in its own publications. Rather, it uses its own subdivision of origin groups and migrant generations, which does not totally overlap with the definitions used in ResSegr.

- Category of origin (Statistics Netherlands definition): the country a person has a connection with, based on the country of birth of either one of the parents or him or herself. Explanation: 'autochtoon' (native Dutch): a person of whom both parents were born in The Netherlands. First generation 'allochtoon': a person born abroad with at least one parent born abroad. Second generation 'allochtoon': a person born in The Netherlands with at least one parent

born abroad. An 'autochtoon' has The Netherlands as category of origin. For a first generation 'allochtoon', the country of birth is his or her category of origin. For a second generation 'allochtoon' the category of origin is the mother's country of birth, unless this is The Netherlands as well. In that case the category of origin is determined by the father's country of birth.

The most important difference is that Statistics Netherlands considers persons with one Dutch and one foreign parent as descendants, whereas these persons are seen as natives in the ResSegr definition. Therefore the countries of birth of person and parents are used to construct the indicators for migrants and foreign origin according to the ResSegr definition.

The definition of a migrant for this indicator is therefore as follows:

- A person is classified as an immigrant if the code for the country of birth is not equal to '6030' (The Netherlands), the code for country of birth of the father is not equal to '6030' and the code for the country of birth of the mother is not equal to '6030'.
- A person is not classified as an immigrant if the code for the country of birth is equal to '6030', or the code for the country of birth is equal to '6030' but the code for the country of birth of the father and/or the mother is equal to '6030'.

The definition for a descendant for this indicator is as follows:

- A person is classified as a descendant if the code for country of birth is equal to '6030' (The Netherlands) but the country of birth of both parents is associated with one of the 28 EU countries (excluding The Netherlands) or one of the four EFTA countries, this person is categorised as an 'EU/EFTA immigrant'.
- A person is considered of foreign origin if he or she can be classified as either an immigrant or a descendant, according to the abovementioned definitions.

The quality of the data from the municipal registers is very high. The data are integral and cover all persons that were or have been included in the municipal registers since 1995. This guarantees a research population that is representative for the Dutch population.

The data are fully comparable from 1995, the first year the statistics were released, onward. There are thus no breaks between the 2003 and 2011 data files (Statistics Netherlands 2017b).

**Belgium – Indicator 2e. Share of the population for persons with foreign origin (immigrants+descendants) in an EU28/EFTA country**

Indicator 2e includes EU28/EFTA immigrants: persons born in an EU28/EFTA country and whose parents were not born in Belgium, and EU28/EFTA descendants: persons born in Belgium whose mother was born in an EU28/EFTA country (or father, if information on the mother is missing). The calculation used individuals' country of birth and kinship. The country of birth of every individual with legal residence in Belgium is known from the 2011 Census variable 'place of birth'. In order to control for the country of births of the parents, this variable was complemented with family relationships contained in the National Register.

There are two potential sources of bias in the construction of the indicators of immigrants and persons with foreign origin for Belgium.

The first one relates to the coverage of the National Register of Natural Persons, from which the population in 2011 is drawn. On the one hand, there is under-declaration of departures, that is, Belgians and foreigners who leave the country and do not report their out-migration to the municipal administration. On the other hand, there is under-declaration of arrivals, both from irregular immigrants, and from European workers living in Belgium who choose not to establish their main residence in the country. However, these coverage issues are expected to have little impact on the population data, since municipalities proceed to regular verification of their records.

The second potential source of bias concerns the parents' origin. If there is no information on the parents' country of birth, a native-born person is not considered as a descendant. As 32% of Belgian-born individuals have both parents of unknown origin, these were not included among the descendants, making up for an underestimation of foreign-origin persons in indicator 2e. Statistics Belgium is currently working on an algorithm to identify parents' origin; once the new data are ready we expect to update this indicator and obtain more accurate measures.

*6.7 Indicator 2f. Share of the population for persons with foreign origin (immigrants+descendent) in an NON-EU28/EFTA country*

**Denmark – Indicator 2f. Share of the population for persons with foreign origin (immigrants+descendent) in an NON-EU28/EFTA country**

An immigrant is defined as a person who is born abroad and whose parents are both (or one in case we only have information of the either the father or mother) foreign citizens or born abroad. Immigrants with parents born in Denmark but who have foreign citizenship are defined as immigrants. If there is no information available about the parents and the person is born in Denmark but is a foreign citizen, he/ she is taken as a descendent. Calculations by Statistics Denmark show that the number of persons with foreign originimmigrants according to this definition is 0.7 % higher compared to the project definition.

A descendant is defined as a person born in Denmark and whose parents (or one of them if only information about the father or mother is available) have foreign citizenship and are immigrants. If we have no information about the parents and the person is born in Denmark and is a foreign citizen, he/ she is taken as a descendant. Calculation based on recent data show that the indicator include 1.6 % extra persons, compared to the project definition for this indicator.

The quality of the classification of persons as immigrants and descendants is generally high. The classification is based on the registration of country of origin and citizenship in the Central Population Register (CPR).

Country of birth is obtained from the CPR run by the ministry of Social and Interior, on January 1 every year. The 98 municipalities register country of birth for all people who immigrate and take residence in the municipality. Municipalities use the register for administrative purposes and have an interest in keeping the quality as high as possible.

Information about the parents in order to define if the person is an immigrant or descendant is available for Statistics Denmark in the population register which source is the CPR. CPR links parents with their (adopted) children.

There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the authorities. Analysis has shown that the population is slightly overestimated (Statistics Denmark (2017a)). For immigrants and descendants, analyses show that the overcoverage is close to 1 %.

Information for each EU/EFTA country is available for all immigrants and descendants. The delimitation of the EU/EFTA group is rather precise.

**Sweden – Indicator 2f. Share of the population for persons with foreign origin (immigrants+descendent) in an NON-EU28/EFTA country**

The Swedish project team only had access to grouped information on country of birth. This meant that a person with foreign origin (immigrants+descendants) in a Non-EU28/EFTA country is defined as: someone who is not born in the EU28 region or one of the Nordic countries, and who does not have a parent born in Sweden, plus those born in Sweden and without parents or a parent born in Sweden, but with parents or a parent not born in the EU28 region or one of the Nordic countries.

This means that the Swedish definition of variable 2f includes immigrants from Switzerland and Lichtenstein. This should however not have any large effects as the number of people registered in Sweden that is born in Switzerland and Lichtenstein in the years 2003 and 2011 is relatively small, see table 1.

Country of birth is obtained from Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the Swedish Tax Agency population register (Folkbokföringsregistret). There is some overcoverage in the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

Information about who are a person's parents is collected through the Swedish multi-generation register (Flergenerationsregistret). The multi-generation register is based on the Swedish Tax Agency population register and is a register of persons who have been registered in Sweden at some time since 1961 and who were born in 1932 or later – so-called index persons. The register contains connections between index persons and their biological parents. For adopted index persons, adoptive parents are registered. The coverage for index persons is almost complete from 1968 onwards. For the period 1961 to 1967, coverage is good, but not as comprehensive (Statistics Sweden 2011b).

If no information exists on the birthplace of the parents the person is considered an immigrant if he/she is foreign-born.

## **Norway – Indicator 2f. Share of the population for persons with foreign origin (immigrants+descendent) in an NON-EU28/EFTA country**

Data on immigrant background is taken from the population register, and provided by Statistics Norway. This register contains information on each individual's immigrant category. Immigrants are defined as people born abroad (meaning that the mother was not registered as resident in Norway at the time of birth), with two foreign-born parents. This definition excludes people born abroad with one or more Norwegian-born parents. Descendants are defined as people born in Norway with two foreign-born parents. This definition excludes people born in Norway with one Norwegian-born parent.

The specific variable used to identify immigrants and descendants in the data is called “invkat”, and immigrants are identified by the value “B”, while descendants are identified by the value “C”.

Immigrants from non-EU/EFTA-countries are defined as immigrants whose mother was a resident in a non-EU/EFTA-country at the time of birth.

Descendants of immigrants from from non-EU/EFTA-countries are defined as descendants whose parents were born in a non-EU/EFTA-country. In cases where the parents have different countries of birth, origin is defined by the mother's country of birth (Statistics Norway 2017c).

For immigrants, information on the mother's country of residence at birth is given by the variable “fodeland”. For descendants, information on parents' country of birth is given by the variable “landbak3gen”. Both of these variables follow the same coding. EU28/EFTA countries are defined by the values 000 (Norway), 101 (Denmark), 103 (Finland), 105 (Iceland), 106 (Sweden), 112 (Belgium), 113 (Bulgaria), 115 (Estonia), 117 (France), 119 (Greece), 121 (Ireland), 122 (Croatia), 123 (Italy), 124 (Latvia), 126 (Malta), 127 (Netherlands), 128 (Liechtenstein), 129 (Luxembourg), 131 (Poland), 132 (Portugal), 133 (Romania), 136 (Lithuania), 137 (Spain), 139 (United Kingdom), 141 (Switzerland), 144 (Germany), 146 (Slovenia), 152 (Hungary), 153 (Austria), 157 (Slovakia), 158 (Czech Republic) and 500 (Cyprus). Note that this means that 102 (Greenland), 104 (Faroe Islands), 114 (Andorra), 118 (Gibraltar), 130 (Monaco), 134 (San Marino) and 154 (Vatican City State) are not included as EU28/EFTA countries.

The data covers everyone registered as residents in Norway. This means that people who reside in Norway without being registered as settled (due to short stays, illegal immigration,

delays in registration, because they await a decision on an application for asylum, or for other reasons) are not counted. Further, people who emigrate without informing the authorities may not be excluded from the registers (Statistics Norway 2017b).

### **The Netherlands – Indicator 2f. Share of the population for persons with foreign origin (immigrants+descendent) in an NON-EU28/EFTA country**

Information on descendants come from the GBA PERSOONTAB dataset and is based on country of birth.

Statistics Netherlands generally does not use the country of birth (of person and parents) indicators to define ‘category of origin’ in its own publications. Rather, it uses its own subdivision of origin groups and migrant generations, which does not totally overlap with the definitions used in ResSegr.

- Category of origin (Statistics Netherlands definition): the country a person has a connection with, based on the country of birth of either one of the parents or him or herself. Explanation: ‘autochtoon’ (native Dutch): a person of whom both parents were born in The Netherlands. First generation ‘allochtoon’: a person born abroad with at least one parent born abroad. Second generation ‘allochtoon’: a person born in The Netherlands with at least one parent born abroad. An ‘autochtoon’ has The Netherlands as category of origin. For a first generation ‘allochtoon’, the country of birth is his or her category of origin. For a second generation ‘allochtoon’ the category of origin is the mother’s country of birth, unless this is The Netherlands as well. In that case the category of origin is determined by the father’s country of birth.

The most important difference is that Statistics Netherlands considers persons with one Dutch and one foreign parent as descendants, whereas these persons are seen as natives in the ResSegr definition. Therefore the countries of birth of person and parents are used to construct the indicators for migrants and foreign origin according to the ResSegr definition.

The definition of a migrant for this indicator is therefore as follows:

- A person is classified as an immigrant if the code for the country of birth is not equal to ‘6030’ (The Netherlands), the code for country of birth of the father is not equal to ‘6030’ and the code for the country of birth of the mother is not equal to ‘6030’.

- A person is not classified as an immigrant if the code for the country of birth is equal to '6030', or the code for the country of birth is equal to '6030' but the code for the country of birth of the father and/or the mother is equal to '6030'.

The definition for a descendant for this indicator is as follows:

- A person is classified as a descendant if the code for country of birth is equal to '6030' (The Netherlands) but the code for the country of birth of both parents is associated with a country other than one of the 28 EU countries (including The Netherlands) or one of the four EFTA countries.

A person is considered of foreign origin if he or she can be assigned to either immigrants or descendants, according to the abovementioned definitions.

The quality of the data from the municipal registers is very high. The data are integral and cover all persons that were or have been included in the municipal registers since 1995. This guarantees a research population that is representative for the Dutch population.

The data are fully comparable from 1995, the first year the statistics were released, onward. There are thus no breaks between the 2003 and 2011 data files (Statistics Netherlands 2017b).

### **Belgium – Indicator 2f. Share of the population for persons with foreign origin (immigrants+descendent) in an NON-EU28/EFTA country**

Indicator 2f includes non-EU28/EFTA immigrants: persons born in a non-EU28/UEFTA country and whose parents were not born in Belgium, and non-EU28/EFTA descendants: persons born in Belgium whose mother was born in a non-EU28/UEFTA country (or father, if information on the mother is missing). The calculation used individuals' country of birth and kinship. The country of birth of every individual with legal residence in Belgium is known from the 2011 Census variable 'place of birth'. In order to control for the country of births of the parents, this variable was complemented with family relationships contained in the National Register.

There are two potential sources of bias in the construction of the indicators of immigrants and persons with foreign origin for Belgium.

The first one relates to the coverage of the National Register of Natural Persons, from which the population in 2011 is drawn. On the one hand, there is under-declaration of departures, that is, Belgians and foreigners who leave the country and do not report their out-migration to the municipal administration. On the other hand, there is under-declaration of arrivals, both from

irregular immigrants, and from European workers living in Belgium who choose not to establish their main residence in the country. However, these coverage issues are expected to have little impact on the population data, since municipalities proceed to regular verification of their records.

The second potential source of bias concerns the parents' origin. If there is no information on the parents' country of birth, a native-born person is not considered as a descendant. As 32% of Belgian-born individuals have both parents of unknown origin, these were not included among the descendants, making up for an underestimation of foreign-origin persons in indicator 2f. Statistics Belgium is currently working on an algorithm to identify parents' origin; once the new data are ready we expect to update this indicator and obtain more accurate measures.

#### *6.8 Indicator 3a. Share of persons 25-64 year olds who have completed tertiary education*

#### **Denmark - Indicator 3a. Share of persons 25-64 year olds who have completed tertiary education**

The information about age is based on Statistics Denmark's register of the total population. The register of the total population is daily updated by events including internal and international migration, births and deaths from the Central Population Register (CPR) which originally is registered by the local registers in the 98 municipalities in Denmark. There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the authorities. Analysis has shown that the population is slightly overestimated (Statistics Denmark (2017a).

Information about completed tertiary education is based on the variable HFAUDD in Statistics Denmark's Longitudinal integration database for educational attainment (PSD\_HOEJSTE\_FULDFOERT\_UDD). HFAUDD describes the highest completed education level and is registered on October 30 every year.

For people with Danish origin, information about the highest completed education is missing for less than 1 % of the population. Information about the highest completed education is missing for all migrants who arrived to Denmark after 2006 unless they have completed an education in Denmark after they have arrived.

### **Sweden - Indicator 3a. Share of persons 25-64 year olds who have completed tertiary education**

The information about age is based on Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the Swedish Tax Agency population register (Folkbokföringregistret). There is some overcoverage in the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

Information about completed tertiary education ("avslutad eftergymnasial utbildning") is based on the variable Sun2000niva in Statistics Sweden's Longitudinal integration database for health insurance and labour market studies (LISA). Sun2000niva describes the highest completed education level and is registered on December 31 (Statistics Sweden 2011a).

In a study from Statistics Sweden (Statistics Sweden 2006), the underestimation of the number of people with tertiary education was calculated to be between 15 and 18 percent. Another report (Statistics Sweden 2014) shows that missing data on highest completed education level is higher for foreign born compared to those born in Sweden. The missing data rate is especially high for foreign born who recently arrived to Sweden. The rate of missing data decreases over time as the chances increases to make a "footprint" in some of the records used in the update of the highest completed education level variable.

### **Norway - Indicator 3a. Share of persons 25-64 year olds who have completed tertiary education**

Age is calculated from birth year, which is available from the population registry data provided by Statistics Norway. Note that we use the age on January 1 of each year.

Information on education is taken from the National Education Database, which draws information from several sources, including registers and census data, and is provided by Statistics Norway. This database is updated annually, and provides detailed information on the education of the Norwegian population, both as annual records of completed education and as panel data on student progression. Analyses conducted on Norwegian data in this project are based on information on each individual's highest attained educational level on October 1 in a given year (Statistics Norway 2017d). This means that people completing a tertiary education

before October 1 in a given year are assigned tertiary education for that year, while people completing tertiary education after October 1 in a given year are assigned non-tertiary education level for that year.

The variable denoting highest attained level of education (called “BU”) is coded by a standard called NUS2000 (see Rognan and Barrabés 2001 for details). People with completed tertiary education are identified by the values 6, 7 and 8, which covers educations at the university or college level. These codes correspond to educational level 5 and 6 in the ISCED97 standard. Note that people with missing information on their educational level are classified as not having completed tertiary education.

### **The Netherlands - Indicator 3a. Share of persons 25-64 year olds who have completed tertiary education**

Data on age has been derived from the municipal population registers (GBA PERSOONTAB), which include the year of birth (Statistics Netherlands 2017b).

Information on Tertiary education has been derived from the file HOOGSTEOPLTAB. As was explained in Chapter 4, HOOGSTEOPLTAB is made up of data from various education registers and a series of volumes of the Survey on the Working Population (EBB, available since 1996). Over the past decades, Statistics Netherlands has gained access the following registers of enrollment and certificates by publicly funded educational institutes:

- Higher education (CRIHO), since 1983 (universities) and 1986 (polytechnics)
- Exam registers for secondary schools (ERR), since 1999
- Education ID codes for secondary education (since 2003/2004), secondary professional education (MBO, since 2004/2005) and adult education (since 2005/2006)
- Data related to study scholarships (WSF), since 1995
- Files with education histories, as listed by unemployed at UWV, since 2010 (only used if no other data available through other registers)

The data contains codes for each study/ education that was finished or followed, including a detailed description. This ‘education-ID’ is a unique 6-digit number in the Standard Education Subdivision (SOI). Based on these codes, the level of education can be derived. A simplified variable for Level of education (achieved) is available, which consists of 3 categories (low,

medium and high education). The high education indicator is comparable with ISCED-codes 5-8 and includes higher vocational and university education.

All persons between 30 and 64 years old who have been assigned to code '3' (high education) have completed tertiary education.

The HOOGSTEOPLTAB database (part of SSD) has a better coverage than previous databases with education data (before 1999). Estimations have become more reliable, especially for small samples. Another advantage is that most data are derived from education registers, which are known for their good quality. The national government uses education registers to control the efficiency of investments in education, and therefore these registers are thoroughly checked by accountants before being published.

Still there are a number of downsides to the education data. Foremost, there are many missing cases; in fact much more than in the other databases that are part of SSD. This is due to the fact that the file only has records for people of whom the highest education followed or finished is known. For many older people no data are available, since education registers have existed only for relatively short period. Furthermore, for many immigrants (reliable) information on education followed abroad before migrating to The Netherlands is largely lacking. Also private education is excluded from the education registers. For these reasons, information from the Survey on the Working Population (EBB) is added to fill gaps where necessary. Thus, the information is integral for a part of the population, but based on samples for other persons.

The quality of the data from the municipal registers (age) is very high. The data are integral and cover all persons that were or have been included in the municipal registers since 1995. This guarantees a research population that is representative for the Dutch population.

The HOOGSTEOPLTAB is available for both 2003 and 2011. In the meantime, the number of data sources has increased, however, which means that the 2011 database is somewhat more extensive than the 2003 one. The following data sources are included in the 2011 version, but were not yet available in 2003.

- Education ID codes for secondary education (since 2003/2004), secondary professional education (MBO, since 2004/2005) and adult education (since 2005/2006)
- Files with education histories, as listed by unemployed at UWV, since 2010 (only used if no other data available through other registers)

This difference has to be taken into account when comparing education data for the two years. The number of missing cases is relatively high for both years, however, compared to other files within the SSD (Statistics Netherlands, 2016a).

### **Belgium - Indicator 3a. Share of persons 25-64 year olds who have completed tertiary education**

As mentioned in chapter 4, information on educational attainment from the Census 2011 is based on the data collected in 2001, updated with the the linguistic communities' databases on the new diplomas obtained since 2001. However, there are potential problems with this indicator. Firstly, persons with missing values in the 2001 Census who have not attained a diploma in 2011 will have missing values in the Census 2011. Secondly, diplomas obtained abroad are very poorly covered. Individuals who attain their diplomas in foreign countries seldom request their recognition in Belgium. Furthermore, only the Flemish register of foreign diplomas was used in in the census variable. Overall, information on education attainment is missing for 9.5 % of individuals older than 15 (see Census results by Statistics Belgium at <http://census2011.fgov.be/censusselecion/selectionFR.html>).

In the light of the problems concerning foreign diplomas, it should be highlighted that the indicator 3b (tertiary education among native-borns) is likely to be more reliable than indicator 3a (tertiary education among the total population).

### *6.9 Indicator 3b. Share of persons among native born 25-64 year olds who have completed tertiary education*

### **Denmark – Indicator 3b. Share of persons among native born 25-64 year olds who have completed tertiary education**

The information about age is based on Statistics Denmark's register of the total population. The register of the total population is daily updated by events including internal and international migration, births and deaths from the Central Population Register (CPR) which originally is registered by the local registers in the 98 municipalities in Denmark. There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the

authorities. Analysis has shown that the population is slightly overestimated (Statistics Denmark (2017a)).

A native born is defined as someone who was born in Denmark. Place (country) of birth is obtained from Statistics Denmark register of the total population published each January 1. The source is the CPR run by the Ministry of Social and Interior. The 98 municipalities register the place/country of birth for all people who register in the local register and take residence in the municipality. The municipality uses the register for administrative purposes and have an interest in keeping the quality as high as possible.

Information about completed tertiary education is based on the variable HFAUDD in Statistics Denmark's Longitudinal integration database for educational attainment (PSD\_HOEJSTE\_FULDFOERT\_UDD). HFAUDD describes the highest completed education level and is registered on October 30 each year.

For people with Danish origin, information about the highest completed education is missing for less than 1 % of the population.

### **Sweden – Indicator 3b. Share of persons among native born 25-64 year olds who have completed tertiary education**

A native born is defined as someone who was born in Sweden. Country of birth is obtained from Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the Swedish Tax Agency population register (Folkbokföringregistret). There is some overcoverage in the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

Information about age is also based on Statistics Sweden's register of the total population (Registret över totalbefolkningen).

Information about completed tertiary education (“avslutad eftergymnasial utbildning”) is based on the variable Sun2000niva in Statistics Sweden's Longitudinal integration database for health insurance and labour market studies (LISA). Sun2000niva describes the highest completed education level and is registered on the 31<sup>st</sup> of December (Statistics Sweden 2011a).

In a study from Statistics Sweden (Statistics Sweden 2006), the underestimation of the number of people with tertiary education was calculated to be between 15 and 18 percent.

### **Norway - 3b. Share of persons among native born 25-64 year olds who have completed tertiary education**

Age is calculated from birth year, which is available from the population registry data provided by Statistics Norway. Note that we use the age on January 1 each year.

Data on immigrant background is taken from the population register, and provided by Statistics Norway. This register contains information on each individual's immigrant background. Native born is defined by one's mother's country of residence at birth, given by the variable "fodeland", where the code 000 denotes Norway.

Information on education is taken from the National Education Database, which draws information from several sources, including registers and census data, and is provided by Statistics Norway. This database is updated annually, and provides detailed information on the education of the Norwegian population, both as annual records of completed education and as panel data on student progression. Analyses conducted on Norwegian data in this project are based on information on each individual's highest attained educational level on October 1 in a given year (Statistics Norway 2017d). This means that people completing a tertiary education before October 1 in a given year are assigned tertiary education for that year, while people completing tertiary education after October 1 in a given year are assigned non-tertiary education level for that year.

The variable denoting highest attained level of education (called "BU") is coded by a standard called NUS2000 (see Rognan and Barrabés 2001 for details). People with completed tertiary education are identified by the values 6, 7 and 8, which covers educations at the university or college level. These codes correspond to educational level 5 and 6 in the ISCED97 standard. Note that people with missing information on their educational level are classified as not having completed tertiary education.

### **The Netherlands – Indicator 3b. Share of persons among native born 25-64 year olds who have completed tertiary education**

The native born in The Netherlands are all persons born in The Netherlands, based on the country of birth in the population registers (GBA PERSOONTAB database). If code '6030' has been assigned to a person for the country of birth variable, this indicates a person is born in The Netherlands. All other codes indicate a person is born abroad.

Data on age has been derived from the municipal population registers (GBA PERSOONTAB), which include the year of birth (Statistics Netherlands 2017b).

Information on Tertiary education has been derived from the file HOOGSTEOPLTAB. As was explained in Chapter 4, HOOGSTEOPLTAB is made up of data from various education registers and a series of volumes of the Survey on the Working Population (EBB, available since 1996). Over the past decades, Statistics Netherlands has gained access the following registers of enrollment and certificates by publicly funded educational institutes:

- Higher education (CRIHO), since 1983 (universities) and 1986 (polytechnics)
- Exam registers for secondary schools (ERR), since 1999
- Education ID codes for secondary education (since 2003/2004), secondary professional education (MBO, since 2004/2005) and adult education (since 2005/2006)
- Data related to study scholarships (WSF), since 1995
- Files with education histories, as listed by unemployed at UWV, since 2010 (only used if no other data available through other registers)

The data contains codes for each study/ education that was finished or followed, including a detailed description. This ‘education-ID’ is a unique 6-digit number in the Standard Education Subdivision (SOI). Based on these codes, the level of education can be derived. A simplified variable for Level of education (achieved) is available, which consists of 3 categories (low, medium and high education). The high education indicator is comparable with ISCED-codes 5-8 and includes higher vocational and university education.

All persons born in The Netherlands between 25 and 64 years old who have been assigned to code ‘3’ (high education) have completed tertiary education.

The HOOGSTEOPLTAB database (part of SSD) has a better coverage than previous databases with education data (before 1999). Estimations have become more reliable, especially for small samples. Another advantage is that most data are derived from education registers, which are known for their good quality. The national government uses education registers to control the efficiency of investments in education, and therefore these registers are thoroughly checked by accountants before being published.

Still there are a number of downsides to the education data. Foremost, there are many missing cases; in fact much more than in the other databases that are part of SSD. This is due to the fact that the file only has records for people of whom the highest education followed or finished is

known. For many older people no data are available, since education registers have existed only for relatively short period. Furthermore, for many immigrants (reliable) information on education followed abroad before migrating to The Netherlands is largely lacking. Also private education is excluded from the education registers. For these reasons, information from the Survey on the Working Population (EBB) is added to fill gaps where necessary. Thus, the information is integral for a part of the population, but based on samples for other persons.

The quality of the data from the municipal registers is very high. The data are integral and cover all persons that were or have been included in the municipal registers since 1995. This guarantees a research population that is representative for the Dutch population.

The HOOGSTEOPLTAB is available for both 2003 and 2011. In the meantime, the number of data sources has increased, however, which means that the 2011 database is somewhat more extensive than the 2003 one. The following data sources are included in the 2011 version, but were not yet available in 2003.

- Education ID codes for secondary education (since 2003/2004), secondary professional education (MBO, since 2004/2005) and adult education (since 2005/2006)
- Files with education histories, as listed by unemployed at UWV, since 2010 (only used if no other data available through other registers)

This difference has to be taken into account when comparing education data for the two years. The number of missing cases is relatively high for both years, however, compared to other files within the SSD (Statistics Netherlands 2016a).

### **Belgium – Indicator 3b. Share of persons among native born 25-64 year olds who have completed tertiary education**

As mentioned in chapter 4, information on educational attainment from the Census 2011 is based on the data collected in 2001, updated with the the linguistic communities' databases on the new diplomas obtained since 2001. However, there are potential problems with this indicator. Firstly, persons with missing values in the 2001 Census who have not attained a diploma until 2011 have also missing values in the Census 2011. Secondly, diplomas obtained abroad are very poorly covered. Individuals who attain their diplomas in foreign countries seldom request their recognition in Belgium. Furthermore, only the Flemish register of foreign diploma was used in in the census variable. Overall, information on education attainment is missing for 9.5 % of

individuals older than 15 (see Census results by Statistics Belgium at <http://census2011.fgov.be/censusselection/selectionFR.html>).

In the light of the problems concerning foreign diplomas, it should be highlighted that the indicator 3b (tertiary education among native-borns) is likely to be more reliable than indicator 3a (tertiary education among the total population).

#### *6.10 Indicator 4a. Share of 25-64 year olds who have a level of taxable earned income in the highest decile*

#### **Denmark – Indicator 4a. Share of 25-64 year olds who have a level of taxable earned income in the highest decile**

The information about age is based on Statistics Denmark's register of the total population. The register of the total population is daily updated by events including internal and international migration, births and deaths from the Central Population Register (CPR) which originally is registered by the local registers in the 98 municipalities in Denmark. There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the authorities. Analysis has shown that the population is slightly overestimated (Statistics Denmark (2017a)).

The income data is based on final tax statements from the tax authorities. Taxable earned income includes wages, received fees subject to labour market contributions and income from self employment. The income statistics register has full population coverage.

#### **Sweden – Indicator 4a. Share of 25-64 year olds who have a level of taxable earned income in the highest decile**

Information about age is based on Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the Swedish Tax Agency population register (Folkbokföringregistret). There is some overcoverage in the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

Taxable earned income is based on the variable ForvInk in Statistics Sweden's Longitudinal integration database for health insurance and labour market studies (LISA). The variable ForvInk is the yearly summary of the gross salary and income from active businesses, provided that the income from active businesses is positive. ForvInk is registered on December 31. Gross salary is reported to the Swedish tax office by the employer. Income from active businesses is based on tax returns (Statistics Sweden 2011a).

The level of taxable earned income in the highest decile is then calculated based on the taxable earned income for all 25-64 year olds for the reference year.

#### **Norway – Indicator 4a. Share of 25-64 year olds who have a level of taxable earned income in the highest decile**

Age is calculated from birth year, which is available from the population registry data provided by Statistics Norway. Note that we use the age on January 1 each year.

Taxable earned income is based on tax return data provided by Statistics Norway. Taxable earned income (“wyrkinnt”) is the sum of wage income and net business income before tax during the calendar year. Sickness benefits and maternity/paternity benefits were also included before 2006 (Statistics Norway 2017e).

The highest decile is calculated from the taxable earned income distribution of people in the age group 25-64 who are registered as residents that year.

#### **The Netherlands – Indicator 4a. Share of 25-64 year olds who have a level of taxable earned income in the highest decile**

Data on age has been derived from the municipal population registers (GBA PERSOONTAB), which include the year of birth (Statistics Netherlands 2017b).

Data on income are derived from the Integral Personal Income (IPI) database, which includes information on personal incomes for all persons registered in The Netherlands, for whom a source of income is known by the national tax office. Besides the actual personal income, also the percentile groups for personal income from the IPI database are available: for the subdivision into percentile groups only persons with personal income have been included. All these persons are subdivided into 100 groups of equal size.

Since the data are derived from tax registers, the quality and reliability is in general very high. The data are integral for all persons that are registered in The Netherlands, which ensures a representative research population.

However, the database does not include all sources of income. Sources that cannot be linked to individual household members, such as rent subsidies and income from capital, are not included in the personal income.

Personal income is not equal to welfare. A single person with only AOW (pensions) earns 70 percent of the amount that couples would earn together. In contrast, standardized household income is equal for both households.

Furthermore, negative incomes are possible (0.5 percent of the population). This is often related to surpluses of self-employed persons or with incomes from other sources.

The quality of the data from the municipal registers (age) is very high. The data are integral and cover all persons that were or have been included in the municipal registers since 1995. This guarantees a research population that is representative for the Dutch population.

There are no breaks in the data; the IPI databases (part of SSD) is measured the same way in both 2003 and 2011 (Statistics Netherlands 2016b).

#### **Belgium – Indicator 4a. Share of 25-64 year olds who have a level of taxable earned income in the highest decile**

Income data in Belgium, based on tax records, are expected to cover all individuals' declared ('legal') economic earnings. The data quality is considered to be good, as the Federal Public Service Finance proceed to consistency checks of tax returns.

The income data covers taxable income only, that is, it does not include non-taxable earnings (social assistance, child benefits, study grants, etc., as well as most income from capital).

Indicator 4a includes, on the other hand, taxable social transfers (such as unemployment and retirement pension).

*6.11 Indicator 5a. Share of persons aged 18-64 year old who received social assistance at some point in the reference year*

**Denmark - Indicator 5a. Share of persons aged 18-64 year old who received social assistance at some point in the reference year**

The information about age is based on Statistics Denmark's register of the total population. The register of the total population is daily updated by events including internal and international migration, births and deaths from the Central Population Register (CPR) which originally is registered by the local registers in the 98 municipalities in Denmark. There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the authorities. Analysis has shown that the population is slightly overestimated (Statistics Denmark (2017a)).

The statistics about social assistance are based on the Cash Benefits register that is administered and maintained by Statistics Denmark. The Cash Benefits register is a full sample based on information from an administrative IT-system in the Danish municipalities called KMD-aktiv. The statistics are based on information regarding the amounts paid to the recipients of cash benefits on an individual-level. The comparability over time is affected by two overall factors. (1) The source of the Cash benefits register was changed to KMD-aktiv in 2007. This meant that the statistics became more precise in the registration of the periods, and in the counting of the recipients. These changes meant a decrease in the number of recipients included in the statistics, but it is not possible to quantify the effect of these changes. (2) The regulation of social assistance is subject to a lot of changes. E.g. the types of benefits and the group of individuals entitled for the specific benefits.

**Sweden - Indicator 5a. Share of persons aged 18-64 year old who received social assistance at some point in the reference year**

Information about age is based on Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the Swedish Tax Agency population register (Folkbokföringregistret). There is some overcoverage in the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

Information about social assistance is based on the variable SocBidrPersF in 2003 and SocBidrPersF04 in 2011, in Statistics Sweden's Longitudinal integration database for health insurance and labour market studies (LISA). The difference between SocBidrPersF and SocBidrPersF04 is that the consumption weights that were used for the calculation were changed after 2003. This however, does not affect indicator 5a, as it only measures whether a person received social benefits, and not how much that person received. SocBidrPersF and SocBidrPersF04 are calculated by taking the household's total social assistance, and multiplying it with the individuals' consumption weights and dividing the sum by the family's total consumption weight. SocBidrPersF and SocBidrPersF04 are registered on December 31 every year. SocBidrPersF and SocBidrPersF04 are based on information provided by all municipalities in Sweden (Statistics Sweden 2011a). Social assistance ("försörjningsstöd", commonly known as "social bidrag"), in Sweden, is a financial support for households with financial problems. The possibility of obtaining social assistance is affected by the whole household's income and assets.

#### **Norway - Indicator 5a. Share of persons aged 18-64 year old who received social assistance at some point in the reference year**

Age is calculated from birth year, which is available from the population registry data provided by Statistics Norway. Note that we use the age on January 1 each year.

The information on social assistance is taken from the income data provided by Statistics Norway. This data is collected from several sources. Further information on how the data was gathered is provided by Steinkellner (2003). The information on receiving social assistance is given by the variable "sosial". This variable provides information on social assistance and social loans from municipal social services (Statistics Norway 2017g). Here we use information on whether or not an individual received any payments in the form of social assistance or social loans from municipal social services in the reference year. Receiving social assistance is defined by having a non-missing value larger than zero on this variable.

#### **The Netherlands - Indicator 5a. Share of persons aged 18-64 year old who received social assistance at some point in the reference year**

Data on age has been derived from the municipal population registers (GBA PERSOONTAB), which include the year of birth (Statistics Netherlands 2017b).

Data on the socioeconomic category are derived from the SECMBUS (Socioeconomic Category per month) database. In order to determine the score on the SECM (socioeconomic category per month) variable, all income sources in a certain month are compared. The highest amount of money determines the socioeconomic category a person belongs to. Also being enrolled at an education institute is taken into account. The following categories are distinguished:

- 11 Employee
- 12 Director-CEO
- 13 Self-employed
- 14 Other active
- 21 Receiver of (temporary) unemployment benefits
- 22 Receiver of welfare benefits [bijstandsuitkering]
- 23 Receiver of other social welfare benefits
- 24 Receiver benefits related to being ill/disabled
- 25 Receiver pensions
- 26 Not yet attending school/ attending school/ studying combined with income
- 31 Not yet attending school/ attending school/ studying without income
- 32 Other without income

The codes 21, 22 and 23 indicate that a person has received social assistance at the end of the reference year. For 2011, we also have these data for each month.

Since the data are derived from tax registers, the quality and reliability is in general very high. The data are integral for all persons that are registered in The Netherlands, which ensures a representative research population.

A note that should be made, however, is that the socioeconomic category is based on income sources, and not all income sources are available in the Social Statistical Database (SSD). Maintenance allowances and income from capital are not included in SSD and therefore not taken into consideration.

For 2011, we also know for each individual category whether a person has belonged to it during a certain month within the reference year (yes =1) or not (no=0). Based on that, it can be derived whether a person received unemployment benefits during the reference year. For 2003, we only have data on the situation at the end of the reference year (Statistics Netherlands, 2017a).

### **Belgium - Indicator 5a. Share of persons aged 18-64 year old who received social assistance at some point in the reference year**

The data concerning persons who received social assistance in Belgium is highly accurate. It relies on register data from the Public Centre for Social Welfare (Centre public d'action sociale/ Openbaar centrum voor maatschappelijk welzijn), centralised in the Crossroad Bank of Social Security (Banque carrefour de la Sécurité sociale/ Kruispuntbank Sociale Zekerheid). It is nevertheless important to highlight that it only concerns individuals who received social assistance in the last month of 2010.

#### *6.12 Indicator 6a. Share of 25-64 year olds in employment*

### **Denmark - Indicator 6a. Share of 25-64 year olds in employment**

The information about age is based on Statistics Denmark's register of the total population. The register of the total population is daily updated by events including internal and international migration, births and deaths from the Central Population Register (CPR) which originally is registered by the local registers in the 98 municipalities in Denmark. There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the authorities. Analysis has shown that the population is slightly overestimated (Statistics Denmark (2017a)).

The information about employment is based on register-based labour force statistics. A person is in employment if he or she was employed for any span of time during the last week of November. Concerning employment, the information after 2008 is generally estimated as being of high quality. Before 2008, the source for employment for employees had a lower quality. As a consequence, the number of employed persons is about 1.7 % lower after 2007. Undeclared work and persons working abroad are not covered by the register-based labour force statistics.

### **Sweden - Indicator 6a. Share of 25-64 year olds in employment**

Information about age is based on Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the

Swedish Tax Agency population register (Folkbokföringregistret). There is some overcoverage in the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

Information about employment is based on the variable SyssStatJ, from Statistics Sweden's Longitudinal integration database for health insurance and labour market studies (LISA). The variable SyssStatJ shows if an individual was employed for any span of time during the month of November in the reference year. This means that variable 6a is defined as the share of 25-64 year olds that were employed sometime during November of the reference year (Statistics Sweden 2011a).

There are a number of coverage problems for the SyssStatJ variable described in the LISA documentation. One is undeclared work, i.e. when the employer has not issued any statement of earnings to the tax agency. Another problem consists of errors that occur when the payment of wages and work is not done during the same year. Yet another example is the group of people who live in Sweden and work abroad, which are classified as "not working" in the register. Overall, those with loose connections to the labour market have the highest risk of being misclassified (Statistics Sweden 2011a).

### **Norway - Indicator 6a. Share of 25-64 year olds in employment**

Age is calculated from birth year, which is available from the population registry data provided by Statistics Norway. Note that we use the age on January 1 each year.

Information on work income is taken from registers maintained by the Norwegian Tax Administration and provided by Statistics Norway. Employment is defined as having a work income in a given year greater than or equal to two basic amounts. The basic amount is a standard unit in the Norwegian welfare and pension systems, and it is adjusted for inflation each year. The basic amounts were 56861 NOK in 2003, 66812 NOK in 2007 and 79216 NOK in 2011. This definition excludes some people working part-time at relatively low wages and/ or people with short unemployment spells during the year.

Information on work income is given by the variable "wyrkinnt", which measures the sum of wage income and net business income during the calendar year. Sickness benefits and maternity/paternity benefits were also included before 2006 (Statistics Norway 2017e).

The reason why we define employment as income above a somewhat arbitrary income threshold is that data on employment for all individuals covering the entire year is not directly available. Other register sources provide only information on a specific week, or for a subsample of the population.

### **The Netherlands - Indicator 6a. Share of 25-64 year olds in employment**

Data on age has been derived from the municipal population registers (GBA PERSOONTAB), which include the year of birth (Statistics Netherlands 2017b).

Data on the socioeconomic category is derived from the SECMBUS (Socioeconomic Category per month) database. In order to determine the score on the SECM (socioeconomic category per month) variable, all income sources in a certain month are compared. The highest amount of money determines the socioeconomic category a person belongs to. Also being enrolled at an education institute is taken into account. The following categories are distinguished:

- 11 Employee
- 12 Director-CEO
- 13 Self-employed
- 14 Other active
- 21 Receiver of (temporary) unemployment benefits
- 22 Receiver of welfare benefits [bijstandsuitkering]
- 23 Receiver of other social welfare benefits
- 24 Receiver benefits related to being ill/disabled
- 25 Receiver pensions
- 26 Not yet attending school/ attending school/ studying combined with income
- 31 Not yet attending school/ attending school/ studying without income
- 32 Other without income

The codes 11, 12, 13 and 14 indicate that a person was employed at the end of the reference year. For 2011, we also have these data for each month.

Since the data are derived from tax registers, the quality and reliability is in in general very high. The data are integral for all persons that are registered in The Netherlands, which ensures a representative research population.

A note that should be made, however, is that the socioeconomic category is based on income sources, and not all income sources are available in the Social Statistical Database (SSD). Maintenance allowances and income from capital are not included in SSD and therefore not taken into consideration.

The quality of the data from the municipal registers (age) is very high. The data are integral and cover all persons that were or have been included in the municipal registers since 1995. This guarantees a research population that is representative for the Dutch population.

For 2011, we also know for each individual category whether a person has belonged to it during a certain month within the reference year (yes =1) or not (no=0). Based on that, it can be derived whether a person received unemployment benefits during the reference year. For 2003, we only have data on the situation at the end of the reference year (Statistics Netherlands 2017a).

### **Belgium - Indicator 6a. Share of 25-64 year olds in employment**

The information on ‘employed persons’ in the Census 2011 is considered to be highly accurate. The primary source, the Crossroad Bank of Social Security (Banque carrefour de la Sécurité sociale/ Kruispuntbank Sociale Zekerheid), gathers information on occupational status from all the relevant institutions in Belgium related to the labour market. In particular, the administrative data from the Census 2011 offers a more reliable picture compared to the traditional censuses until 2001, in which individuals were asked about their status (Eurostat 2014).

#### *6.13 Indicator 6b. Share of 30-59 year old in employment*

### **Denmark - Indicator 6b. Share of 30-59 year old in employment**

The information about age is based on Statistics Denmark’s register of the total population. The register of the total population is daily updated by events including internal and international migration, births and deaths from the Central Population Register (CPR) which originally is registered by the local registers in the 98 municipalities in Denmark. There is some overcoverage in the register of 2011 because some people who emigrate do not report their emigration to the

authorities. Analysis has shown that the population is slightly overestimated (Statistics Denmark (2017a)).

The information about employment is based on register-based labour force statistics. A person is in employment if he or she was employed for any span of time during the last week of November. Concerning employment, the information after 2008 is generally estimated as being of high quality. Before 2008, the source for employment for employees had a lower quality. As a consequence, the number of employed persons is about 1.7 % lower after 2007. Undeclared work and persons working abroad are not covered by the register-based labour force statistics.

### **Sweden - Indicator 6b. Share of 30-59 year old in employment**

Information about age is based on Statistics Sweden's register of the total population (Registret över totalbefolkningen). The register of the total population is an extract from the Swedish Tax Agency population register (Folkbokföringregistret). There is some overcoverage in the register of the total population, i.e. some people are registered as living in Sweden while they actually have emigrated or died (Statistics Sweden 2015b).

Information about employment is based on the variable SyssStatJ, from Statistics Sweden's Longitudinal integration database for health insurance and labour market studies (LISA). The variable SyssStatJ shows if an individual was employed for any span of time during the month of November in the reference year. This means that variable 6a is defined as the share of 30-59 year olds that were employed sometime during November of the reference year (Statistics Sweden 2011a).

There are a number of coverage problems for the SyssStatJ variable described in the LISA documentation. One is undeclared work, i.e. when the employer has not issued any statement of earnings to the tax agency. Another problem consists of errors that occur when the payment of wages and work is not done during the same year. Yet another example is the group of people who live in Sweden and work abroad, these are classified as "not working" in the register. Overall, those with loose connections to the labour market have the highest risk of being misclassified (Statistics Sweden 2011a).

### **Norway - Indicator 6b. Share of 30-59 year old in employment**

Age is calculated from birth year, which is available from the population registry data provided by Statistics Norway. Note that we use the age on January 1 each year.

Information on work income is taken from registers maintained by the Norwegian Tax Administration and provided by Statistics Norway. Employment is defined as having a work income in a given year greater than or equal to two basic amounts. The basic amount is a standard unit in the Norwegian welfare and pension systems, and it is adjusted for inflation each year. The basic amounts were 56861 NOK in 2003, 66812 NOK in 2007 and 79216 NOK in 2011. This definition excludes some people working part-time at relatively low wages and/or people with short employment spells during the year.

Information on work income is given by the variable “wyrkinnt”, which measures the sum of wage income and net business income during the calendar year. Sickness benefits and maternity/paternity benefits were also included before 2006 (Statistics Norway 2017e).

The reason why we define employment as income above a somewhat arbitrary income threshold is that data on employment for all individuals covering the entire year is not directly available. Other register sources provide only information on a specific week, or for a subsample of the population.

### **The Netherlands - Indicator 6b. Share of 30-59 year old in employment**

Data on age has been derived from the municipal population registers (GBA PERSOONTAB), which include the year of birth (Statistics Netherlands 2017b).

Data on the socioeconomic category is derived from the SECMBUS (Socioeconomic Category per month) database. In order to determine the score on the SECM (socioeconomic category per month) variable, all income sources in a certain month are compared. The highest amount of money determines the socioeconomic category a person belongs to. Also being enrolled at an education institute is taken into account. The following categories are distinguished:

- 11 Employee
- 12 Director-CEO
- 13 Self-employed
- 14 Other active
- 21 Receiver of (temporary) unemployment benefits

- 22 Receiver of welfare benefits [bijstandsuitkering]
- 23 Receiver of other social welfare benefits
- 24 Receiver benefits related to being ill/disabled
- 25 Receiver pensions
- 26 Not yet attending school/ attending school/ studying combined with income
- 31 Not yet attending school/ attending school/ studying without income
- 32 Other without income

The codes 11, 12, 13 and 14 indicate that a person was employed at the end of the reference year. For 2011, we also have these data for each month.

Since the data are derived from tax registers, the quality and reliability is in in general very high. The data are integral for all persons that are registered in The Netherlands, which ensures a representative research population.

A note that should be made, however, is that the socioeconomic category is based on income sources, and not all income sources are available in the Social Statistical Database (SSD). Maintenance allowances and income from capital are not included in SSD and therefore not taken into consideration.

The quality of the data from the municipal registers (age) is very high. The data are integral and cover all persons that were or have been included in the municipal registers since 1995. This guarantees a research population that is representative for the Dutch population.

For 2011, we also know for each individual category whether a person has belonged to it during a certain month within the reference year (yes =1) or not (no=0). Based on that, it can be derived whether a person received unemployment benefits during the reference year. For 2003, we only have data on the situation at the end of the reference year (Statistics Netherlands 2017a).

### **Belgium - Indicator 6b. Share of 30-59 year old in employment**

The information on ‘employed persons’ in the Census 2011 is considered to be highly accurate. The primary source, the Crossroad Bank of Social Security (Banque carrefour de la Sécurité sociale/ Kruispuntbank Sociale Zekerheid), gathers information on occupational status from all the relevant institutions in Belgium related to the labour market. In particular, the administrative data from the Census 2011 offer a more reliable picture compared to the traditional censuses until 2001, in which individuals were asked about their status (Eurostat 2014).

## 7. Summary of differences between variables and countries for different years

### 7.1 General remarks about the data in the five countries

Most of the data gathered for the ResSegr indicators rely on administrative sources. These register data have the advantage of covering the quasi-totality of the countries' residents and they offer more accurate measures compared to traditional surveys or censuses. However, it is also important to acknowledge some inherent limitations of administrative sources, common to the five countries.

The official population figures provided by register-based sources may somewhat differ from the *de facto* population. On the one hand, there is under-registration of departures, i.e. persons who leave the country and do not report their emigration to the local authorities. On the other hand, many individuals remain 'invisible' to administrative sources, namely irregular migrants, and regular European/ Scandinavian migrants who choose not to establish their official domicile in their country of residence. Moreover, some individuals may be registered in a dwelling and reside in fact in another place (e.g. students).

Likewise, administrative sources do not cover irregular employment arrangements and undeclared sources of income. Only the 'formal' economy is apparent in the sources.

Despite these limitations, register-based data remain undoubtedly the most reliable sources at our disposal, especially in the five ResSegr countries which have a long tradition of administrative data collection and management. In fact, Statistic Offices and local administrations in the five countries proceed to regular verifications and statistical tests in order to provide accurate demographic and socioeconomic figures.

### 7.2 Geographic information

Grid sizes are strictly comparable among Belgium, Denmark, the Netherlands, and Norway: each country was divided into 100m x 100m cells. The only exception is that analyses from Norway that cover the entire country are based on a 400m x 400 m grid for the k-values 25600 and 51200 for reasons explained in section 5.3 of this report. In contrast, Swedish grids are larger: 250m x 250m in urban areas, and 1,000m x 1,000m in rural areas. The Swedish particularity must be taken into account in comparative studies among the five countries.

As far as the projections are concerned, the x and y coordinates correspond to the lower left-hand corner of the grid cell in Denmark, the Netherlands and Norway; and corresponds to the centroids of the grids in Belgium and Sweden. These differences in coordinates do not affect the international comparison of segregation patterns.

### *7.3 Immigrants and descendants*

As mentioned above, it is important to keep in mind that the register data used in the ResSegr project do not cover irregular migrants. In spite of this, the data about immigrants in the five countries provide an accurate account of the registered foreign-born population.

As for descendants, data is considered to be of good quality in all countries except for Belgium. In fact, there is a poor coverage of descendants in this country, as the parents' country of birth is missing for 32 % of Belgium-born individuals.

Moreover, the following particularities in Belgium and Sweden must be highlighted:

- The Belgian data is the only one to include refugees waiting for asylum.
- In Sweden, immigrants and descendants from Switzerland and Liechtenstein are counted among the 'non-EU/EFTA' groups (although their number is fairly limited).

Overall, indicators 2a to 2c can be assuredly compared among the five countries. In contrast, indicators 2d to 2f (which include descendants) are fairly comparable among Denmark, the Netherlands, Norway and Sweden, whereas they should be used with caution in the case of Belgium.

### *7.4 Tertiary education*

All five countries employed an identical concept to construct the indicators of tertiary education. These indicators are based on data about individuals' 'highest education level successfully attained', from which the proportion of people with a tertiary education is derived.

Overall, there is a very good coverage of the education level of native-borns in all countries. However, the registration of diplomas obtained abroad is not equally good. This concerns mostly the immigrant population. The coverage of foreign diplomas is particularly poor in Belgium and the Netherlands, and less so in Scandinavian countries that use surveys or other sources to gather such data (Norway and Denmark).

Considering the potential inconsistencies with the coverage of foreign diplomas, indicator 3b

(tertiary education among native-borns) should yield more accurate measures in comparative studies among the ResSegr countries.

### *7.5 Income*

Income data in the five countries mainly rely on tax registers, which are, in most cases, linked to other administrative databases. Overall, data quality is considered to be good, although it partly depends on each country's efficacy in controlling the consistency of income declarations, errors in tax returns, undeclared income, errors, etc.

The income data is used for the calculation of two indicators: the share of persons at risk of poverty, i.e. with income below 60% of the median level (1a); and the share of persons with high income, defined as those in highest national decile (4a). These indicators rely on different types of income. Whereas indicator 1a is based the concept of 'disposable income after social transfer', the second one only takes into account the taxable earned income from wages or net-earnings from self-employment (see chapter 3).

There are nevertheless important differences in the income data among the five countries, which necessarily affect the way disposable income and taxable earned income are obtained in each case. The most significant discrepancies are the following:

- In Sweden, income data is available at the individual level, and a consumption weight is used in order to estimate individual income.
- In Norway, personal disposable income (indicator 1a) is defined as the sum of work income, capital income, taxable transfers and tax-free transfers, minus taxes and negative transfers, while taxable earned income (indicator 4a) is defined as work income before tax.
- The Dutch income data are measured as income after paid income insurances but before tax, and also do not include rent subsidies and income from capital.
- In Belgium, no distinction could be made between disposable income (for indicator 1a) and taxable earned income (for indicator 4a). The only income data available refers to taxable income. These do not include non-taxable earnings (social assistance, child benefits, study grants, etc., as well as most income from capital). Therefore, both indicators 1a and 4a were calculated based on taxable income.

As a consequence of these differences in data availability, whereas indicators 1a and 4a give a general idea, respectively, of the spatial distribution of poverty and wealth, they are not strictly

comparable among the five countries.

In sum, indicators 1a and 4a, based on income data, are not strictly comparable among countries nor over time. They can provide an interesting idea of economic segregation *patterns* in the countries and of how economic segregation relates to the other indicators of socioeconomic and ethnic segregation. However, one must be cautious in comparing segregation *levels* among countries based on indicators 1a and 4a.

### 7.6 *Social assistance*

Social assistance systems are not identically organised in the five countries. In fact, for each country, the indicator ‘social assistance’ (5a) may include different things (for example, full-time unemployed in Denmark, any public benefits granted to households in Sweden, social assistance payments and social loans in Norway etc. — see chapter 6). Moreover, the reference period can vary from one country to the other, from any time in a given year (as in the case of Norway and the Netherlands in 2011) to only a month or point in time in the reference year (Belgium, Sweden, the Netherlands in 2003). There are also breaks in the indicator’s time series Norway and, to some lesser, in the Netherlands, which makes comparison in time less accurate in these two countries.

Because of the fundamental differences both in the content of social benefits in the five countries and in the length of reference periods, the social assistance indicator 5a is not strictly comparable among countries. It can give an idea of the spatial distribution — and concentration — of poverty in the countries and its relation to the other indicators, but it should not be used to compare poverty levels.

### 7.7 *Employment*

Employment data (indicators 6a and 6b) present some differences between the five countries.

A first difference between countries is the length of the reference period from which employment figures are drawn: the number of employed persons may refer to the situation at the end of the year (the Netherlands in 2003 and Belgium), in the month of November (Denmark and Sweden) or at any point in the year (the Netherlands in 2011). These differences, however, should not have a major impact on the comparability of employment indicators.

The most significant difference concerns the Norwegian data. Whereas the four other

countries use a similar method to produce employment figures, based on administrative databases of the labour force (employed, self-employed, temporarily absent, etc.), in Norway employed persons are identified based on individual's earnings from work in the reference year. Therefore, employment indicators in Norway may be not strictly comparable to the other countries.

The comparability of employment indicators over time may be biased in the case of Denmark, where significant changes in employment registration took place between 2003 and 2011. In the case of the Netherlands, employment data is expected to be more accurate for 2011 in respect to 2003 (see chapter 6).

## 8. Ethical considerations and privacy

### 8.1. Denmark

Danish data produced within ResSegr are subjected to Statistics Denmark's rules governing data confidentiality (Statistical Code of Statistics Denmark) as described on the website of Statistics Denmark: <http://www.dst.dk/en/OmDS/lovgivning>. It appears from the rules that the non-disclosure practice also covers "aggregated statistical products for publication" (page 3). Output from the EquiPop model is summed data and thereby aggregated data.

Under the item "Protection of confidentiality in social statistics" (p. 7) it is stated that "Information in statistical products, which can be assigned to individual persons, must be subjected to non-disclosure." and further "It must not be possible to extract new knowledge about the circumstances of individual persons from a statistical table. The non-disclosure practice is applied to such an extent, which ensures that this is not possible."

In this respect, a potential conflict may arise in relation to the aggregated summed data at the level of individualized neighborhood. On the basis of an estimated RATIO of 0 or 100, it can subsequently be derived that none/all in the neighborhood in question (sum of grid cells) or in the grid cell (100x100 metres) forming the basis of the enumeration in the neighborhood in question, fulfils/does not fulfil the indicator (see description of the 13 indicators). The geographic identification of the grid cell forming the basis of the enumeration of neighborhood can be performed by means of the coordinates, and thereby the individual persons in the grid cell can be identified.

In connection with a RATIO of 0, respectively, 100 %, the non-disclosure problem is equal, irrespective of whether there is 1 or 100 persons within the grid cell, as they are characterized by fulfilling/ or not fulfilling the indicator (for example, characterized by indicator 5a "Share of persons aged 18-64 year old who received social assistance at some point in the reference year"). For this reason, we cap the data, i.e. the grids with shares of 0 to 0.05 are replaced with <0.05 and grids with a share of 0.95 to 1 are replaced with >0.95 (similar to Sweden). Consequently, it will no longer be possible to obtain a homogeneous identification, and consequently compliance with Statistics Denmark's data confidentiality policy is achieved.

Information at the level of individuals is not sold nor delivered by Statistics Denmark and Statistics Denmark reserves its right not to disclose data to such an extent, which is required by the non-disclosure practice of information with regard to persons, families, households and business enterprises.

## 8.2. Sweden

The Swedish register data are administered in the GeoStar database, which is a collection of databases managed at Statistics Sweden, and accessed via the MONA system (<http://www.scb.se/mona-en/>).

The terms and conditions under which these data can be used within ResSegr are based on three documents: a confidentiality agreement, Statistics Sweden's general terms of agreement, and the ethical review that has been conducted.

The confidentiality agreement contains regulations on who may use these data and under which conditions. It states that the data can only be used within the project that was approved after ethical review. One clause states that material may only be published in a way so that single individuals or business identities are not disclosed.

The general terms of agreement contain, among others, regulations that prohibit the sale or otherwise commercial exploitation of material. It also includes a clause stating that personal data should be treated within the framework of the personal data act.

Research conducted within ResSegr falls within the project "Geografisk kontext: Ett nytt sätt att mäta vad omgivningen betyder för individens livsbana" (Geographical context: A new way to measure what context means for individual life courses), that was approved by the Stockholm Regional Ethical Vetting Board (Regionala Etikprövningsnämnden i Stockholm). The data does not reveal individual identities. The project description describes that geographical coordinates will be used to construct contextual variables for differently sized individualized neighbourhoods and to link contextual data to individual data, but that coordinates will not be used to identify individuals. Data are kept by Statistics Sweden on a server that project members can connect to. Identifiable data may under no circumstances be exported from Statistics Sweden.

The regulations for use of EquiPop include the prohibition to profit from the use of the software, for instance users may not sell research reports, presentations and other forms of output

and analyses that were produced via EquiPop (<http://equipop.kultgeog.uu.se/Legal/untitled.html>). Statistics Sweden has not allowed the use of the EquiPop software within the MONA system. Based on the approved ethical review, users of the GeoStar data have been granted the possibility to export aggregated data from the MONA system to their own computers, where they then can use EquiPop to process data. The data is aggregated on grid level, meaning that individuals cannot be identified. Statistics Sweden has agreed that such data may be published.

The only possibility to disclose information on individuals in Sweden in the harmonized multi-country datasets is when grids have a ratio of either 0 or 1 for an indicator described in chapter 6. For instance in the case of the share of persons in employment, this would mean that either no or all individuals in a grid are in employed. For this reason, we cap the data, i.e. the grids with shares of 0 to 0.05 are replaced with <0.05 and grids with a share of 0.95 to 1 are replaced with >0.95.

### 8.3. Norway

Non-health-related research conducted on Norwegian register data requires approval from the Norwegian Data Protection Authority. The Norwegian contribution to this project was carried out with permission from the Norwegian Centre for Research Data (NSD) which functions as the Data Protection Official for the University of Oslo, and The Norwegian Data Protection Authority. These authorities deemed that the potential benefits from the project were sufficient to allow the use of detailed register data on anonymous individuals without attaining informed consent.

All data has been made available by Statistics Norway, and directly identifying information has been removed from the provided files. Data for this project has been kept on a secure server maintained by the University of Oslo. According to the Norwegian statistics act, personal data may not be transferred out of the country. Therefore, all analyses on individual data is performed on the servers of the University of Oslo, and only aggregate non-sensitive statistics are shared with our collaborators abroad.

The restrictions on the use of the program EquiPop for profit (<http://equipop.kultgeog.uu.se/Legal/untitled.html>) are adhered to in this project. The preparation of data using EquiPop was done on a secure server maintained by the University of Oslo.

With regards to map data and detailed grid-level data, the researchers on the Norwegian project team are currently (as of October 10, 2017) working on determining the amount of detail on neighbourhood characteristics that we are allowed to share through maps and data. The level of detail provided from analyses of the Norwegian data will depend on the outcome of this process.

#### *8.4. Netherlands*

A first distinction should be made between data that are to be made available for the researchers within the project and data for the general public. The general rule at Statistics Netherlands is that individual data cannot be shared with the general public. Individual researchers within the project are allowed to work on the aggregated data, as long as clear bilateral agreements are made. This is generally no problem for researchers within the EU. This way, Statistics Netherlands keeps control over who has access to the data. At relatively short term, there will be a cooperation between the Dutch and Belgian research teams, which requires members of both research teams to have access to aggregated output for these two countries.

It is, however, possible to share frequency tables and maps with the general public, as long as they comply to the following rules:

- Tables must consist of aggregated information and not information at the individual level.
- There is a minimum number of observations: all tables and similar output should have at least 10 units as a basis for each cell or data point. This applies both to persons and companies. In case tables consist of percentages, also the original numbers should be mentioned. Minimum and maximum scores are not released, since they generally refer to one unit.
- Regarding the issue of group disclosure, in each frequency table or similar output no single cell may contain more than 90% of the total number of units in each row or column. This is in order to prevent that certain variables in a table form an identifiable group, where other variables in the table reveal information that is valid for each group member. Although no individual unit can be recognized, confidentiality is violated since information is valid for each member of the group, and the group in total can be identified.

- 'Flat' maps can be presented to the general public, including the interactive webpage. However, it is not allowed to link underlying data to these maps; for example when clicking on a specific datapoint leads to a map with detailed information.

These rules are especially strict when it comes to 'sensitive' indicators, such as income and social assistance. At all time, Statistics Netherlands has the right to check output before it leaves the Statistics Netherlands environment.

We will comply with these regulations in the following way. First, although the 100x100m grids are the core of each k-level, aggregated statistics are calculated and presented only for the k-levels, thus for a minimum of 200 persons, and not for the grids or individuals within them. This decreases chances of identification, since many grids have a very small population count.

Second, since for specific indicators there are cells with only a very small number of people, presented output will be restricted to ratios and distance and no absolute numbers are presented. For each k-level, the number of persons belonging to a subgroup will not be delivered. Only the ratio, and the total population of the k-level are presented. This partly solves the problem that cells with less than 10 persons may not be presented.

Third, very low or very high percentages are truncated: scores from 0 to 10 percent are taken together in the category 'less than 10 percent' and scores of 90 to 100 percent are grouped in the category 'more than 90 percent'.

The restrictions on the use of the program EquiPop for profit (<http://equipop.kultgeog.uu.se/Legal/untitled.html>) are adhered to in this project. The EquiPop analyses were conducted on a secured stand-alone computer, maintained by Statistics Netherlands.

### *8.5. Belgium*

Most of the data used in the construction of the ResSegr indicators for Belgium rely on the Census 2011 (except for income data, which were supplied in addition to the Census). This is the first census carried out in the country that was entirely based on the linkage between different administrative sources, such as the National Register and the Crossroad databank of social security. The population in 2011 was subsequently linked to the cadastre databank in 2011, making it possible to locate the place of residence of the census population using geo-

coordinates. For the ResSegr project, Statistics Belgium operated an aggregation of the geolocated population within grids of 100m x 100m for the entire Belgian territory.

Statistics Belgium owns the rights of the Census 2011 data. According to the Belgian law, Statistics Belgium is allowed to publish and disseminate ‘global and anonymous figures’ (i.e. aggregate data), unless the small number of individuals in a given statistic unit makes it possible to identify individual information (see article 2 in [http://www.ejustice.just.fgov.be/cgi\\_loi/change\\_lg.pl?language=fr&la=F&cn=1962070430&table\\_name=loi](http://www.ejustice.just.fgov.be/cgi_loi/change_lg.pl?language=fr&la=F&cn=1962070430&table_name=loi)).

In the case of the EquiPop outputs and geocoded data in general, the problem of small numbers often arises. Moreover, a number of specific/isolated locations in the territory are considered to be sensitive. For this reason, Statistics Belgium has analysed the content of EquiPop outputs for each one of the indicators in order to assure that the final datasets do not disclose any individual or sensitive information. In particular, the total and minority populations inside the grids (CountAllLocal and CountAllGroup columns) were not made available to the researchers.

In practice, it was agreed with Statistics Belgium that only the researchers within the ResSegr project will have access to EquiPop files, excluding the columns CountAllLocal (population inside the grids) and CountGroupLocal (minority group inside the grids). The complete EquiPop files are still available at Statistics Belgium, and analyses that need these two columns (e.g. calculation of segregation indices) can be carried out on on-site computers; the final aggregate results of the analyses will be checked by Statistics Belgium before leaving the site.

Concerning the dissemination outside the ResSegr project, namely in the ResSegr webpage, only non-interactive ‘flat’ maps may be used (as in the case of the Netherlands). In other words, the underlying data of these maps (ratios linked to coordinates) will not be made available to the general public.

## References

- Bakker, Bart F., Van Rooijen, Johan and Van Toor, Leo (2014). The system of social statistical datasets of Statistics Netherlands: An integral approach to the production of register-based social statistics. *Statistical Journal of the IAOS* 30(4), 411-424.
- Chaix, Basile, Merlo, Juan, Subramanian, S.V., Lynch, John and Chauvin, Pierre (2005). Comparison of a spatial perspective with the multilevel analytical approach in neighbourhood studies: The case of mental and behavioural disorders due to psychoactive substance use in Malmö, Sweden, 2001. *American Journal of Epidemiology* 162(2), 171-182.
- Chaix, Basile, Merlo, Juan, Evans, David, Leal, Cinira and Havard, Sabrina (2009). Neighbourhoods in eco-epidemiologic research: delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur. *Social Science & Medicine* 69(9), 1306-1310.
- Clark, William, Andersson, Eva, Östh, John and Malmberg, Bo (2015). A multi-scalar analysis of neighbourhood composition in Los Angeles 2000-2010: a location based approach to segregation and diversity. *Annals of the Association of American Geographers*, doi: 10.1080/00045608.2015.1072790.
- Duncan, Otis Dudley and Duncan, Beverly (1955). Residential distribution and occupational stratification. In C. Peach (Ed.), *Urban Social Segregation* (pp. 51-67). London: Longman.
- Eurostat (2014). 2011 Census Hub: Textual metadata. <https://ec.europa.eu/CensusHub2/metadata.do> [accessed on April 1, 2017].
- Fisher, Claude S., Stockmayer, Gretchen, Stiles, Jon and Hout, Michael (2004). Distinguishing the geographical levels and social dimensions of U.S. metropolitan segregation, 1960-2000. *Demography* 41(1), 37-59.
- Fowler, Christopher S. (2015). Segregation as a multiscale phenomenon and its implications for neighbourhood-scale research: the case of South Seattle 1990–2010. *Urban Geography*, (ahead-of-print), 1-25.
- Finney, Nissa and Catney, Gemma(2012). *Minority Internal Migration in Europe*. Ashgate, Farnham, Surrey.
- Lee, Barrett A., Reardon, Sean F., Firebaugh, Glenn, Farrell, Chad R., Matthews, Stephen A. and O’Sullivan, David (2008). Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review* 73, 766-691.

- Lichter, Daniel T., Parisi, Domenico and Taquino, Michael C. (2012). The geography of exclusion: Race, segregation, and concentrated poverty. *Social Problems* 59(3), 364-388.
- Lichter, Daniel T., Parisi, Domenico and Taquino, Michael C. (2015). Toward a new macro-segregation? Decomposing segregation within and between metropolitan cities and suburbs. *American Sociological Review* 80(4), 843-873.
- Lloyd, Christopher D. and Shuttleworth, Ian G. (2009). Are Northern Ireland's communities dividing? Evidence from geographically consistent population data 1971-2001. *Environment and Planning A* 41, 213-229.
- MacAllister, Iain, Johnston, R.J., Pattie, C.J., Tunstall, H., Dorling, D.F.L., and Rossiter, D.J. (2001). Class dealignment and the neighbourhood effect: Miller revisited. *British Journal of Political Science* 31(1), 41-59.
- Malmberg, Bo, Andersson, Eva and Östh, John (2011). To what extent does the level of segregation vary between different urban areas? Introducing a scalable measure of segregation. European Network for Housing Research Conference, 5-8 July, Toulouse. <<http://www.enhr2011.com/sites/default/files/Paper-BoMalmberg-WS16.pdf>>.
- Malmberg, Bo, Andersson, Eva and Östh, John (2013). Segregation and urban unrest in Sweden. *Urban Geography* 34(7), 1031-1046.
- Massey, Douglas S. and Denton, Nancy A. (1988). The dimensions of residential segregation. *Social Forces* 67(2), 281-315.
- Morrill, R.L. (1991). On the measure of geographic segregation. *Geography Research Forum* 11, 25-36.
- Olzak, Susan, Shanahan, Suzanne and McEneaney, Elizabeth H. (1996). Poverty, segregation, and race riots: 1960 to 1993. *American Sociological Review* 61(4), 590-613.
- Openshaw, Stan (1984). The Modifiable Areal Unit Problem. Concepts and techniques in modern geography, Vol. 38. Norwich: Geobooks.
- Östh, John, (2014). Introducing the EquiPop software: An application for the calculation of k-nearest neighbour contexts/ neighbourhoods. <http://equipop.kultgeog.uu.se/Tutorial/Introducing%20EquiPop.pdf>
- Östh, John, Malmberg, Bo and Andersson, Eva (2014). Analysing segregation with individualized neighbourhoods defined by population size. In Lloyd, Christopher D.,

- Shuttleworth, Ian G. & Wong, David W. (Eds.), *Social-Spatial Segregation: Concepts, Processes and Outcomes* (pp. 135-161): Policy Press.
- Östh, John, Clark, William A.V. and Malmberg, Bo (2015). Measuring the scale of segregation using k-nearest neighbor aggregates. *Geographical Analysis*, 47, 34-49.
- Reardon, Sean F., Matthews, Stephen A., O'Sullivan, David, Lee, Barrett A., Firebaugh, Glenn, Farrell, Chad R. and Bischoff, Kendra (2008). The geographic scale of metropolitan racial segregation. *Demography* 45(3), 489-514.
- Reardon, Sean F., Farrell, Chad R., Matthews, Stephen A., O'Sullivan, David, Bischoff, Kendra and Firebaugh, Glenn (2009). Race and space in the 1990s: Changes in the geographic scale of racial residential segregation, 1990–2000. *Social Science Research* 38(1):55–70.
- Rognan, Arild and Barrabés, Nuria (2001). NUS2000 Dokumentasjonsrapport [NUS2000 Documentation report]. Notater 2001/12. Avdeling for personstatistikk/ Seksjon for befolknings- og utdanningsstatistikk. Statistics Norway. [https://www.ssb.no/a/publikasjoner/pdf/notat\\_200112/notat\\_200112.pdf](https://www.ssb.no/a/publikasjoner/pdf/notat_200112/notat_200112.pdf) [accessed October 4 2017].
- Schierup, Carl-Ulrik and Ålund, Aleksandra (2011). The end of Swedish exceptionalism? Citizenship, neoliberalism and the politics of exclusion. *Race & Class* 53(1), 45-64.
- Shuttleworth, Ian G. and Östh, John (2013). Making international segregation comparisons: Does it matter which spatial building blocks are used to construct k-nearest neighbourhoods? Annual Meeting of the American Association of Geographers, 9-13 April, Los Angeles.
- Statistics Belgium (2014). Census 2011. <http://census2011.fgov.be> [accessed April 1, 2017].
- Statistics Denmark (2017a). Præcision og pålidelighed [Precision and reliability]. <http://www.dst.dk/da/Statistik/dokumentation/statistikdokumentation/befolkningen/praecision-og-paalidelighed> [accessed October 23, 2017].
- Statistics Denmark (2017b). Indvandrere og efterkommere [Immigrants and descendants]. <http://www.dst.dk/da/Statistik/dokumentation/statistikdokumentation/indvandrere-og-efterkommere> [accessed October 23, 2017].
- Statistics Netherlands (2014). Documentatierapport Gemeentelijke Basis Administratie adressen van personen [Documentation report Municipal Basic Administration of personal addresses] (GBAADRESBUS). The Hague/ Heerlen: Statistics Netherlands. Downloadable at:

<https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata>.

Statistics Netherlands (2016a). Documentatierapport Hoogst behaald en hoogst gevolgd opleidingsniveau en opleidingsrichting van de bevolking in Nederland [Documentation report Highest education degree and specialization achieved and highest education degree and specialization followed of the Dutch population] (HOOGSTEOPLTAB). The Hague/ Heerlen: Statistics Netherlands. Downloadable at: <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata>.

Statistics Netherlands (2016b). Documentatierapport Integraal Persoonlijk Inkomen (IPI) [Documentation report Integral Personal Income] (IPI). The Hague/ Heerlen: Statistics Netherlands. Downloadable at: <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata>.

Statistics Netherlands (2016c). Documentatierapport X- en Y-coördinaten van een verblijfsobject [Documentation report X and Y coordinates of each object used for accomodation] (VSLCOORDTAB). The Hague/ Heerlen: Statistics Netherlands. Downloadable at: <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata>.

Statistics Netherlands (2017a). Documentatie Personen sociaaleconomische categorie [Documentation Persons socioeconomic category] (SECMBUS). The Hague/ Heerlen: Statistics Netherlands. Downloadable at: <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata>.

Statistics Netherlands (2017b). Documentatierapport Persoonskenmerken van alle in de Gemeentelijke Basis Administratie (GBA) ingeschreven personen, gecoördineerd [Documentation report Personal characteristics of all persons registered in the Municipal Basic Administration, coordinated] (GBAPERSONTAB). The Hague/ Heerlen: Statistics Netherlands. Downloadable at: <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata>.

Statistics Netherlands (2017c). Documentatierapport Adreskenmerken van personen die in de gemeentelijke bevolkingsregisters ingeschreven (hebben ge)staan [Documentation report Adress characteristics of persons who have (or had) been enrolled in the municipal population registers] (GBAADRESOBJECTBUS). The Hague/ Heerlen: Statistics

Netherlands. Downloadable at: <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata>.

Statistics Norway (2014). Data til forskning - steg for steg. [Data till research – step for step].

<https://www.ssb.no/omssb/tjenester-og-verktoy/data-til-forskning> [accessed October 3 2017].

Statistics Norway (2017a). Total income.

<http://www.ssb.no/a/metadatas/conceptvariable/vardok/570/en> [accessed October 3, 2017].

Statistics Norway (2017b). Immigrants and Norwegian-born to immigrant parents.

<http://www.ssb.no/en/befolkning/statistikker/innvbef> [accessed October 3, 2017].

Statistics Norway (2017c). Variabeldefinisjoner. Landbakgrunn. [Definition of variables: Foreign

origin. <http://www.ssb.no/a/metadatas/definisjoner/variabler/likittittel/landbakgrunn.shtml>

[accessed October 3, 2017].

Statistics Norway (2017d). Dokumentasjon. Nasjonal utdanningsdatabase [Documentation on the

national education database]. <http://www.ssb.no/omssb/tjenester-og-verktoy/data-til-forskning/utdanning/nasjonal-utdanningsdatabase> [accessed October 4, 2017].

Statistics Norway (2017e). Variabeldefinisjon. Yrkesinntekter [Definition of variables:

Professional income]. <http://www.ssb.no/a/metadatas/conceptvariable/vardok/49/nb> [accessed

October 6, 2017].

Statistics Norway (2017f). Variabeldefinisjon. Inntekt etter skatt [Definition of variables: Income

after tax]. <http://www.ssb.no/a/metadatas/conceptvariable/vardok/2/nb> [accessed October 6,

2017].

Statistics Norway (2017g). Variabeldefinisjon. Sosialhjelp [Definition of variables: social

assistance]. <http://www.ssb.no/a/metadatas/conceptvariable/vardok/668/nb> [accessed October

10, 2017].

Statistics Sweden (2006). Evalvering av utbildningsregistret [Evaluation of the Swedish register

of education]. Bakgrunnsfakta, Befolknings- och välfärdsstatistik 2006:4. Örebro: Statistics

Sweden.

Statistics Sweden (2008). Ungdomars flytt hemifrån [Adolescents home-leaving]. Demografiska

rapporter 2008:5. Örebro: Statistics Sweden.

Statistics Sweden (2011a). Longitudinell integrationsdatabas för Sjukförsäkrings- och

Arbetsmarknadsstudier (LISA) 1990–2009 [Integrated database for labour market research].

- Bakgrundsfakta, Arbetsmarknads- och utbildningsstatistik 2011:4. Stockholm: Statistics Sweden.
- Statistics Sweden (2011b). SCB:s data för forskning – De mest använda registern [Microdata at Statistics Sweden for research purposes]. Stockholm: Statistics Sweden.
- Statistics Sweden (2014). Utbildningsbakgrund bland utrikes födda [Educational background among foreign born persons]. Tema utbildning. Temarapport 2014:6. Örebro: Statistics Sweden.
- Statistics Sweden (2015a). Tätorter; arealer, befolkning 2010 [Urban areas, areas, population 2010]. [http://www.scb.se/Statistik/MI/MI0810/\\_dokument/MI0810\\_DO\\_2010.pdf](http://www.scb.se/Statistik/MI/MI0810/_dokument/MI0810_DO_2010.pdf).
- Statistics Sweden (2015b). Övertäckning i registret över totalbefolkningen – en registerstudie [Overcoverage in the Total Population Register – a register study]. Befolkning och Välfärd. Bakgrundsfakta 2015:1. Örebro: Statistics Sweden.
- Statistics Sweden (2017) Befolkning efter födelseland, ålder och kön. År 2000-2016 [Population by country of birth, age and gender, 2000-2016]. Statistics Sweden. [http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START\\_BE\\_BE0101\\_BE0101E/FodelslandArK/?rxid=f45f90b6-7345-4877-ba25-9b43e6c6e299](http://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START_BE_BE0101_BE0101E/FodelslandArK/?rxid=f45f90b6-7345-4877-ba25-9b43e6c6e299).
- Steinkellner, Alice (2003) Inntektsstatistikk for personer og familier 1999-2001. Dokumentasjon av datagrunnlag og produksjonsprosess [Income statistics for persons and families 1999-2001. Documentation of the database and the production process]. Notater 2003/72. Avdeling for næringsstatistikk/Seksjon for inntekts- og lønnsstatistikk. Statistics Norway. [http://www.ssb.no/a/publikasjoner/pdf/notat\\_200372/notat\\_200372.pdf](http://www.ssb.no/a/publikasjoner/pdf/notat_200372/notat_200372.pdf) [accessed October 4, 2017].
- Strand Geir-Harald and Bloch, Vilni Verner Holst (2009) Statistical grids for Norway. Documentation of national grids for analysis and visualisation of spatial data in Norway. Documents 2009/9. Statistics Norway/ Department of Economic Statistics. [http://www.ssb.no/a/english/publikasjoner/pdf/doc\\_200909\\_en/doc\\_200909\\_en.pdf](http://www.ssb.no/a/english/publikasjoner/pdf/doc_200909_en/doc_200909_en.pdf) [accessed October 3, 2017].
- The Swedish Tax Agency (2006). Kartläggning av folkbokföringsfelet [Mapping of errors in national registration]. Solna: Skatteverket, analysenheten.
- Wong, David W.S. (2004). Comparing traditional and spatial segregation measures: A spatial scale perspective. *Urban Geography* 25(1), 66-82.